REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
05-10-2013

2. REPORT TYPE

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE

Estimating the Effects of Pre-College Education on College Performance

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kotlikoff, Phoebe McDaniel

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Naval Academy
Annapolis, MD 21402

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Trident Scholar Report no. 417 (2013)

12. DISTRIBUTION / AVAILABILITY STATEMENT

This document has been approved for public release; its distribution is UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This paper assesses the effects of post-secondary education on college success by examining a large detailed cross-sectional dataset of students from the U.S. Naval Academy. We find that students who have attended a pre-college program tend to graduate at higher rates than comparable students entering directly from high school but perform at lower levels academically overall.

15. SUBJECT TERMS
Remediation, Education, Pre-College, Selection bias

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		81	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

U.S.N.A. --- Trident Scholar project report; no. 417 (2013)

ESTIMATING THE EFFECTS OF PRE-COLLEGE EDUCATION ON COLLEGE PERFORMANCE

by

Midshipman 1/c Phoebe M. Kotlikoff
United States Naval Academy
Annapolis, Maryland

____
(signature)

Certification of Advisers Approval

Associate Professor Ahmed S. Rahman
Economics Department

____
(signature)
____
(date)

Associate Professor Katherine A. Smith
Economics Department

____
(signature)
____
(date)

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder
Associate Director of Midshipman Research

____
(signature)
____
(date)

Abstract

This paper assesses the effects of post-secondary education on college success by examining a large detailed cross-sectional dataset of students from the U.S. Naval Academy. We find that students who have attended a pre-college program tend to graduate at higher rates than comparable students entering directly from high school but perform at lower levels academically overall.

Keywords: Remediation, Education, Pre-College, Selection bias

Acknowledgements

This year-long research project is the undertaking of many people and represents a combined effort. I would like to thank the Trident Committee members for the opportunity to undertake this research project. I would also like to recognize the patience, kindness, dedication, and effort put forth by my two advisers, Professor Ahmed S. Rahman and Professor Katherine A. Smith, as well as the entire Economics department for supporting me in this venture.

Table of Contents

I. Introduction

Education policy makers have historically debated the effectiveness of preparatory school programs as a method of post-high school, pre-college education. However, the difficult nature of measuring returns to education, combined with limited access to relevant data, makes coming to quantitative conclusions about the utility of preparatory school programs challenging. This paper assesses the effects of post-secondary education on college success through propensity score matching and instrumental variable analysis by using a large detailed cross-sectional dataset of students from the United States Naval Academy (USNA).

College graduates in the United States tend to have significantly higher earnings and higher labor force participation rates than their counterparts who have only high school degrees. U.S. workers with a bachelor's degree or better will earn an average of $40,000 more in one year than those with a high school degree, illustrating the importance of degree attainment for future earnings (Carnevale et al. 2010). Moreover, according to Census Bureau 2013 employment projections, the unemployment rate for college graduates aged 25 and over is 4.5% compared to 8.3% for high school graduates. Despite the rising importance of college education to future income, there is growing concern that many college entrants are unprepared to succeed in undergraduate studies, contributing to the 42% drop-out rate among students pursuing their bachelor's degrees (Alliance for Excellent Education 2011).

Although record numbers of U.S. high school students are entering college, four in ten are academically unprepared for college-level studies. One method of providing additional preparation to students is to enroll them in pre-college programs that act as an intermediary institution between high school and college. However, these remedial programs are expensive, costing over $3 billion in 2011 (Alliance for Excellent Education 2011). This paper examines the impact of participation in pre-college remedial education on student preparation for undergraduate success, controlling for intrinsic ability.

This analysis is performed using data from USNA. We use a large cross-sectional dataset of students from the USNA graduating classes of 1988 – 2011, categorizing students into those who have attended the Naval Academy Preparatory School (NAPS), an independent pre-college program selected by USNA known as a Foundation school, or another college prior to USNA, and those entering directly from high school. On average, over 20% of each graduating class attended some type of pre-college program, and another 3% were admitted from colleges around the country. Using a combination of propensity score matching and instrumental variable analysis to mitigate selection bias, we determine the empirical relationships between assignment to a post-secondary education program and future success in college. Because NAPS and Foundation school programs are comparable to other pre-college programs in terms of curriculum and environment, the results of this analysis can be extrapolated to apply to pre-college programs nationally, helping to inform U.S. education policy more broadly.

Due to non-randomness in selection of students into pre-college education programs, traditional economic evaluation techniques are ineffective in investigating educational returns. The USNA Admissions Board assesses each candidate's background characteristics, and designates those students who will start at a pre-college program. The main characteristics, considered by the board are Math and Verbal SAT scores and high school class rank. However, there is a broad range of acceptance based on each background characteristic, and each individual is considered independently. This leads to enormous overlap in many background characteristics between the treatment and control groups.

For this reason, we turn to propensity score matching and instrumental variables to eliminate the selection bias associated with assigning students to pre-college education programs. Results are determined by comparing individuals with similar background characteristics in order to identify returns to a number of performance variable estimation including graduation rates, academic course grades, and class rank. The data is split into treatment and control groups. The control observations are individuals who matriculate into USNA directly from high school. The treatment groups include those who attended NAPS, a Foundation school, or another college before USNA.

There are a number of benefits to analyzing only students from USNA. United States college completion rates have declined nationally, due both to changes in the preparedness of entering students and in collegiate characteristics. Analyzing students from USNA allows for a focus on an institution whose characteristics have remained fairly consistent over time, in order to isolate the effects of student preparedness on college success. Moreover, USNA maintains robust and detailed records on each midshipman, including high school information, SAT/ACT scores, attendance records over all four years, and grades during each semester. This degree of detailed accounting of student characteristics and performance is not typical of standard private institutions of higher education. Therefore, this study accounts for biases inherent in other studies in order to isolate the true effects of remedial preparatory programs. Our access to more detailed midshipmen background and performance records facilitates the use of methods like propensity score matching, which mitigate potential misinterpretation caused by selection bias. Our detailed dataset allows for a comparison of treated and untreated groups. Further, midshipmen at USNA are kept in a virtual test-tube environment. The liberty policies, uniform regulations, drug-tolerance policy, and military requirements provide controls for many variables that would be otherwise uncontrollable and unknowable in alternative environments. In addition, first year classes are nearly always identical from student to student due to required coursework. This type of comparison is not an option for studies using data from other colleges. However, USNA is comparable to any top tier civilian undergraduate institution. It is a fully accredited college with a standard curriculum.

The NAPS program is a USNA-run preparatory school designed to ease students into the rigors of higher education. Students who are not admitted to USNA directly from high school may be offered an appointment to NAPS, a tuition-free preparatory school program. Students at NAPS have Navy Reserve status and attend pre-college or college level courses during the day for an entire academic year before matriculation into USNA. The Foundation school program is sponsored by the Naval Academy Foundation. Students not directly admitted to USNA may also be considered for a slot at a Foundation school. Foundation schools include a wide range of military and non-military preparatory schools around the country. Foundation school students are sponsored by the Naval Academy Foundation on a need-based system but are asked to pay at least 40% of tuition.

The results of this study will not only identify the specific returns to investment in NAPS and Foundation schools but will also assess the effectiveness of post-secondary education on a national level. The NAPS program is comparable to pre-college programs that send students to colleges and universities all over the U.S. The Foundation school program places students in independent preparatory school programs alongside students who will go on to attend a wide range of civilian undergraduate institutions. Conclusions about the returns to the NAPS and Foundation school programs can, therefore, be applied to pre-college programs on a national level.

This paper is organized into the following sections: an overview of relevant literature, an explanation of methodology, a description of the data, and a discussion of the empirical results.

II. Background Literature

In this section we discuss the decline of college graduation rates, current evidence for returns to remediation, the evolution of remedial education programs, and finally the rise of pre-college programs to better prepare students for college.

U.S. college completion rates are declining, while the wage gap between those who earn an undergraduate degree and those who do not is widening. The higher wage premium on a college degree has led to an increase in the percentage of students entering college, but there has not been a proportional increase in graduation rates. Bound et al. (2009) hypothesize that both lack of student preparation and deteriorating college characteristics contribute to declining graduation rates. On one hand, the increasing wage premium encourages higher numbers of more weakly prepared students to begin degree programs, but many subsequently drop out. Of the bottom quartile of the student sample tested, the probability of a student attending college jumped from 21.7 to 44.0 percent over 20 years. At the same time, only 5% graduated with a bachelor's degree over the same time period, indicating that more underprepared students are beginning college, but the percent actually attaining degrees remains static. Moreover, the study finds that much of the decline is a result of supply-side changes to institutional characteristics like student-teacher ratio and per-student funding. This is supported by the fact that while college graduation rates have declined, high school graduation rates have stayed fairly constant over the past several decades. Moreover, out of the bottom quartile of college attendees in the dataset, graduation rates fell from 25.8% to 11.4% over two decades. This suggests that of those students on the margin who do decide to attend college, factors other than preparation are also contributing to declining completion rates. The assumptions of the Bound et. al. (2009) paper to initiate an inquiry into the effectiveness of pre-college remedial programs.

Literature on this topic uses the term “remedial” to describe coursework that students should have mastered before entering college. Previous literature has investigated the effectiveness of remedial coursework for students already enrolled in college. According to a study by Attewell, Lavin, Domina, and Levey (2006), remediation programs for students during their first year in college decrease the likelihood of graduating on time by 6%. This follows from the fact that if students are enrolled in remedial programs during their first year, they are falling behind in coursework that their fellow students are taking, and falling behind the four-year schedule of required coursework necessary to graduate on time. Adleman (1999) also suggests an inverse relationship between participation in remedial coursework and subsequent graduation. However, once the data was adjusted to include a measure for high school preparation, the disparity in graduation rates disappeared. Lavin, Alba, and Silberstein (1981) attempt to control for background characteristics through a study on remedial coursework at CUNY that was assigned but not mandatory. They found positive returns from remediation in the form of a slight increase in probability of graduation. Although this study represented a step toward addressing the selection issue, it neglected to account for the fact that students who would voluntarily participate in a remediation program were likely to be more vested in their educational success.

Bettinger and Long (2005) analyze the implications of remedial coursework in mathematics on persistence in attaining a bachelor’s degree. The results controlled for background skills and came to three significant conclusions. First, the students placed in remedial courses in math and English were less likely to drop out of their four-year colleges or transfer to two-year colleges. Second, placement in remediation programs did not lower overall likelihood of attaining a bachelor’s degree. Finally, those students placed in remedial programs who completed those programs were more likely to attain a bachelor’s degree than students who did not complete remedial coursework but were otherwise similar. There was also weak evidence that suggested students in remediation would go on to achieve better grades in their first college-level math courses. The Bettinger and Long study made a complicated comparison between students taking very different classes during their first year of college. In this study, USNA students take very similar courses during the first year of study, allowing for a more direct comparison of academic performance. The literature on this subject suggests that a preparatory year of remedial coursework might dramatically improve on-time graduation rates since it leads to the conclusion that if students are taking a year to prepare themselves for college work, they will be able to follow the expected schedule of courses and graduate from an undergraduate institution in the expected four years with a degree.

There are many options for addressing the problem of student under preparedness for undergraduate education. Many colleges assess student preparedness during the admissions process and then assign them to remedial classes during their freshman year, but this is not the only option for unprepared students. Increasing emphasis has been put on the importance of community colleges and pre-college programs rather than remedial courses in college. Yet little is known about the effectiveness of such programs. This paper contributes quantitative evidence about the returns of pre-college remedial programs, not remedial courses taken as a college freshman. Soliday (2002) suggests that remediation can help students acquire critical skills and enable them to integrate themselves into college populations. Pre-college programs offer an interim year for students who graduate from high school with college aspirations but lack the experience in high-level coursework and the self-discipline to succeed as freshmen students in an undergraduate environment. Preparatory schools like NAPS and Foundation schools give students a taste of the independence and rigor of college, while simultaneously improving their preparedness academically. However, at an annual cost of over one billion dollars for U.S. public colleges alone, critics of remediation wonder if such programs should be offered at all **(**Breneman and Haarlow, 1997).

Thus, this paper examines in what ways pre-college programs are helping better prepare students for the rigors of higher education at the undergraduate level. The preparatory school programs that feed into USNA are designed to prepare students not only for the academic challenges that await them, but also for the high-stress lifestyle of a busy college student. A postgraduate (PG) preparatory year has become popular for many students who are academically unprepared or not competitive for selective four-year undergraduate programs. There are currently 144 schools nationwide offering a PG year for high school graduates (Boarding School Review, http://www.boardingschoolreview.com/). Pre-college programs are designed to better prepare students for the academic rigors of college life. Therefore, participation in a pre-college program should align with some measureable improvement in college readiness, whether in terms of academic success or graduation rates.

A particular problem with current literature on this topic is its inability to hone in on the cohort of students who fall on the margin of acceptance to high-level undergraduate programs. Literature documents a trend in which students from less affluent families tend to need remedial education at higher rates and also drop out at higher rates than more affluent students (Attewell et. al, 2006). While affluence is a proxy for many other background characteristics (education level, health, future earnings), it is important to note that while students are attending USNA, tuition costs and personal living expenses do not play a role in the probability that they will drop out. Consequently, the framework of USNA provides a unique perspective on the issue of cost. As midshipmen, students are not responsible for paying tuition and are required to graduate in four years. In this study, we are then able to eliminate much of the variance associated with examining the performance of students at the margin. At other colleges, a solid sampling of marginal students is difficult to find because many will drop out for a variety of reasons. However, at USNA family income plays a far less significant role in pushing students to drop out of school due to cost or other factors, allowing us to examine the educational returns to a cohort previously difficult to isolate.

Moreover, the literature justifies the extrapolation of conclusions drawn from service academy data to the larger population of undergraduate students. According to extensive research, particularly that of Carrell and West[1], the use of data specific to USNA in undertaking this analysis does not limit the scope of the results to the same population. Although the average Midshipman has many more restrictions and mandatory obligations than the average college student, the results of this study have implications that apply to all types of remedial education outside of the walls of USNA. Moreover, the environment of USNA is controlled in ways described above, which create a dataset particularly suited to this type of longitudinal policy study.

The importance of this study lies in the fact that the controversy over the effectiveness of pre-college remedial programs has been “sporadic, underfunded, and inconclusive” (Merisotis and Phipps 2000). Moreover, many of the existing studies only add to the controversy, either due to conclusions based on qualitative rather than quantitative analysis, inability to obtain a control group, or failure to address the issue of selection bias. In this paper, we take a hard quantitative look at treated and untreated groups using statistical analysis tools that control for selection bias.

---

[1] For example, see Scott E. Carrell & Teny Maghakian & James E. West (2011), Carrell, Scott E. & Hoekstra, Mark & West, James E. (2011), Carrell, Scott E. & Hoekstra, Mark & West, James E. (2011), Scott E. Carrell & James E. West (2010), Scott E. Carrell & Richard L. Fullerton & James E. West (2009), Lyle, David S. (2007).

Moreover, rather than studying remedial coursework in the college setting, we focus on pre-college programs that serve to bring students up to a level at which they should be prepared to start college level coursework.

In summary, previous literature does not adequately investigate the role of pre-college remediation programs. This is due, in part, to the lack of a standard around which all pre-college education methods can be assessed. There is no method of quality assessment across remediation programs. Without assessment, there is no way to determine best practices or analyze programs comparatively. This, in turn, perpetuates a lack of assessment (Bettinger and Long, 2004). Moreover, many studies employ a flawed methodology in their assessment of pre-college utility, comparing treatment and control groups without controlling for selection bias (Bettinger and Long, 2004). This paper applies directly to the debate by examining the specific cohort of students for whom remediation is pertinent. The complex USNA admissions policies create an environment in which many students who are assigned to a pre-college program have very similar background characteristics to students who are admitted directly to USNA. This paper isolates that overlap in characteristics, exploiting the detailed and robust Midshipmen performance records, to analyze the utility of pre-college to the students who might actually benefit from it, rather than the entire student body. Therefore, this paper explores a new niche in assessing whether pre-college is a useful tool in preparing students to succeed in the college environment, and ultimately become degree-holding graduates.

III. Methodology

The general objective of this type of study is to compare differences in outcomes between "treated" and "non-treated" individuals. In this case outcomes are measures of educational achievement, and treatment is enrollment in a pre-college program. Specifically, the untreated group of students in the study is the direct admissions group, and the treatment group is split into three smaller cohorts: NAPS students, Foundation school students, and prior college students.

In order to ensure robust conclusions, this paper utilizes three methods to assess returns from pre-college programs: ordinary least squares (OLS) regression, propensity score matching (PSM), and instrumental variable (IV) analysis. First, we use ordinary least squares regression as a basic method by which to determine the relationship between participation in a pre-college program and certain performance metrics. Next, we generate a propensity score for each individual in our dataset. A propensity score is an individual's probability of attending a given pre-college program based on his/her background characteristics. We can then algorithmically match treated and untreated individuals with similar propensity scores and calculate the differences in their performance at USNA in a technique called propensity score matching. The propensity score proves to be a useful tool for instrumental variable analysis as well. In the third section of this study, we use the propensity score as an instrument in a regression to measure the estimated differences in performance between those who participate in a pre-college program and those who do not.

Ideally, a study would compare the outcomes of those who receive treatment with the outcomes of those same individuals if they were not given the treatment. However, this is an experimental impossibility. Due to the constraints of an observational study, we must find a way to simulate experimental conditions. In this study, that means looking at individuals with similar background characteristics who either attended a pre-college program or did not attend a pre-college program. For this reason, this study uses observed data with propensity score matching and instrumental variable analysis techniques to compare the differences in outcomes between a treated group and a control group. Propensity score matching methods of analysis are particularly useful for assessing the overlap between a treated and control group.

In this study, let *Y* be the outcome variable – this includes a number of options: graduation rates, first year academic GPA, first year aptitude scores, choice of major, overall order of merit, and so forth (each is considered in separate specifications). Let *T* be the treatment. Specifically, $T = 1$ indicates a member of the treated group (in this case those who attend NAPS, or in separate specifications those who attend a Foundation school or prior college), and $T = 0$ indicates a member of the control group (in this case those who enter USNA directly from high school).

The goal is to estimate the mean impact on the measured outcome variable from the treatment, obtained by averaging the impact across all the individuals in the population. *ATE* is the average treatment effect. In general, it is modeled using this equation:

$$ATE = E\big(Y|(T = 1) - Y|(T = 0)\big) \tag{1}$$

where $E(\bullet)$ represents the average or expected value of the outcome. In this study, one specification of this model is:

$$ATE = E\big(G|(N = 1) - G|(N = 0)\big) \tag{2}$$

where G is the probability of an individual graduating, and N is a binary variable documenting that an individual either participated in NAPS (N = 1) or did not participate in NAPS (N = 0). The average treatment effect is a measure of the returns to education between the treated and control groups without considering selection bias. If participants are randomly assigned to treatment, the average difference in outcomes between the treated and control groups is a measure of the impact of the treatment. However, in this case treatment is intentionally non-random – selection for pre-college or remedial programs is based on individual characteristics, some of which are perhaps unobservable, and these characteristics very likely also impact the outcome.

Rather than picking students at random for remediation, students who fulfill program specific criteria are offered assignment to one of the described feeder programs. The non-randomness of the process means the results suffer from a rather severe *selection bias*. Assuming X is a matrix of co-variates capturing student characteristics that potentially affect student outcomes, consider estimating the following linear specification:

$$Y_{it} = \alpha + \beta * T_{it} + \delta * X_{it} + \varepsilon_{it} \qquad (3)$$

β is intended to capture the average treatment effect. However, given the selection bias, the OLS estimate of β will be inconsistent because $E[\varepsilon|T] \neq 0$. For example, in this study, one estimate analyzes the effect of attending NAPS on a student's overall academic of merit at graduation:

$$AOM_{it} = \alpha + \beta * (NAPS_{it}) + \delta_1 * \mathrm{MathSAT_{it}} + \delta_2 * \mathrm{VerbalSAT_{it}} + \varepsilon_{it} \qquad (4)$$

where the error term includes the fact that those students who are given the treatment of attending NAPS are selected for treatment based on their background characteristics. An OLS estimate that includes the background characteristics above will take into account the broad range of performance between treatment and control groups. However, without estimating the matched cohort only, the error term will still be significantly correlated with a given treatment. This study eliminates the demonstrated selection bias by employing propensity score matching techniques. In essence, these techniques use the information from the covariates of those from the control group to observe what would happen to treated individuals if they had in fact not participated in the treatment. By comparing how outcomes differ for treated individuals relative to observationally similar non-treated individuals, it is possible to estimate the effects of the treatment program.

As previously mentioned, many studies on this subject (NCES, 1996; Drosinos, 2004; Fitzpatrick, 2001) utilize simple ordinary least squares regression, or some similar variation, to assess returns from pre-college education programs. OLS regression does have some factors of control in terms of comparing students with similar backgrounds. However, it also includes the entire dataset in analysis, including those students who would never be assigned to pre-college and those who would rarely be accepted before assignment to some type of remedial program. Here lies the benefit of propensity score techniques: propensity score matching addresses the issue of self-selection and allows a decomposition of treatment effects on outcomes through a number of very detailed metrics: average treatment effect (ATE), effect of treatment on the treated (TT), and the potential effect of treatment on the untreated (TUT) (Heckman et al. 2010).

Given that the selection process for treatment is not perfectly known, the first step is to *estimate* the likelihood of treatment based on the background characteristics of the students. This entails estimating a “propensity score” for each student.

$$Propensity\ Score\ (PS) = E(N|X) \tag{5}$$

In this case, N is a binary variable that reflects participation in the NAPS program. We also examined the propensity score for the treated group when treatment is F for the participation variable for Foundation students, and C for prior college students. The variable X refers to a matrix of background characteristics described above: Math and Verbal SAT scores, rank in high school graduating class, and measure of high school quality. The conversion of background information that affects the treatment selection process into a single scalar variable helps alleviate the “curse of dimensionality” originally described by Rosenbaum and Ruben (1983).

There are a number of ways to estimate and use the propensity score to help combat selection bias and thus capture the true effects of treatment. In this study, the primary method of analysis was propensity score matching (Heinrich et al. 2010). Propensity score matching methods are designed to ensure that impact estimates of treatment are based on outcome differences between comparable individuals. This approach exploits the entire sample by employing matching algorithms that use the estimated propensity scores to match untreated individuals to treated ones.

There are two assumptions that must be satisfied in order to implement propensity score matching: the Conditional Independence Assumption (CIA) and the Common Support Condition (CSC). The CIA assumption states that there is a matrix of background characteristics, X, which directly impacts selection into the treatment or non-treatment group. However, once this matrix of characteristics is controlled for, the treatment assignment is, according to Heinrich, comparable to random assignment. This assumption acknowledges the presence of selection bias but also defines the ability to reduce that bias by controlling for differences between groups.

The CSC assumption states that for any value of X, the probability of being treated and being untreated is between 0% and 100%. In other words, there is a positive probability of being both treated and untreated. This assumption sets up the idea of overlap in background characteristics between the treated and untreated groups. In order to “match” scores, there must be observations in which the background characteristics are similar. This similarity is dependent on the existence of overlap between the background variables of both the treated and untreated groups.

The first step is to estimate the propensity score, typically using a logit or probit function:

$$E[T_{it}|X_{it}] = S(X_{it}) \tag{6}$$

where $S(X_{it})$ describes how covariates influence the decision over whether or not to give treatment. This logit or probit function converts background information from many variables into a single measure of the expected likelihood of a person receiving treatment.

This leads into a discussion of the matrix X_{it}, the background characteristics on which students are matched. A quantitative study like this relies on large amounts of descriptive data that can be manipulated to even the playing field between those who attended a pre-college program and those who did not. However, literature stresses the importance of being parsimonious in terms of variable inclusion in the first stage of analysis (Heckman et al. 2010). When creating propensity scores in the first stage, we must only include variables that have a direct impact on both selection to treatment and the outcome variable of interest in the specification. This eliminates certain descriptive statistics from our dataset. For example, while gender and ethnicity are important background characteristics of an individual, they have no explicit impact on an outcome measure like academic performance in the first semester of undergraduate studies. Although they may indirectly affect performance due to their correlation with variables like wealth and quality of previous education, gender and ethnicity variables will increase variance. For this reason, it is essential to select first-stage matching variables that describe only information pertinent to the examined outcome variables. A more in-depth discussion of first-stage model specifications is located in the empirical results section.

After creating a propensity score for each individual, the next step is to match up individuals with "similar" propensity scores from treated and control groups. There are a number of matching algorithms we can employ, the most common of which are nearest neighbor matching, radius matching, and kernel and local-linear matching. Nearest neighbor matching is one of the most straightforward matching procedures. An individual from the control group is matched with an individual from the treatment group in terms of the closest propensity score. One can vary this approach to include matching with or without replacement where, in the former case, a member from the control group can be used more than once as a match (this is potentially important for us as far more students enter the Academy directly as opposed to going through NAPS or a Foundation school) (Caliendo and Kopeinig, 2005). In this study, we employed nearest neighbor matching on the closest nearest neighbor, 5 nearest neighbors, and 20 nearest neighbors, all with replacement. To avoid the potential risk of poor matching, radius matching specifies a maximum propensity score distance (sometimes called a "caliper") by which a match can be made. This approach differs in that it uses not only the nearest neighbor, but also all of the comparison group members within the caliper. That is, it uses as many comparison cases as are available within the caliper, but not those that are poor matches based on the specified distance (Caliendo and Kopeinig, 2005). Finally, kernel and local linear matching are non-parametric matching estimators that compare the outcome of each treated person to a weighted average of the outcomes of all those in the control group. The highest weight is placed on those with scores closest to the treated individuals (Caliendo and Kopeinig, 2005).

While there is no clear rule for determining which algorithm is most appropriate, a key consideration is that algorithm selection involves a clear bias/efficiency tradeoff. Nearest neighbor matching minimizes bias by using only the most similar observations but ignores a lot of information, while local linear matching produces more efficiency but increases the bias by potentially using poorer matches. Although there are many more matching algorithms, these are the methods that balance both sides of the bias/efficiency trade off. They are also the methods that we used to analyze the data in this study (Caliendo and Kopeinig, 2005).

Estimating the returns to pre-college education from programs like NAPS, Foundation schools, and prior college is suited to the technique of propensity score matching by virtue of the amount of data available and the end goal: measuring the effect of a treatment on outcome in an observational scenario in which there is inherent selection bias. It was essential to find a way of measuring treatment effects while taking into account the fact that treatment selection is nonrandom and dependent on a range of variables in designing this study. Moreover, there is no value for any background characteristic for which the probability of treatment is either perfectly zero or perfectly one. This overlap between groups creates a scenario ideal for the use of propensity score matching, while also eliminating more standard methods of analysis due to selection bias.

Another method for eliminating selection bias in treatment versus control studies is to use an instrumental variable. In this study, we used the propensity score itself as an instrumental variable in a second two-stage analysis. In the first stage, we created the propensity score, $\widehat{NAPS_i}$, an estimate of the likelihood of an individual attending NAPS based on a matrix of background characteristics. In the hopes of creating a useful instrumental variable, we selected a very inclusive first-stage equation to calculate the propensity score. It is critical to include variables that affect selection to NAPS but do not directly effect outcome variables like academic order of merit, graduation rates, or academic grades. Our model had to not only include the background characteristics we selected for our propensity score matching model but also information on an individual's ethnicity, home of record, gender, and any other variables that would affect the probability of being assigned to a pre-college program.[2]

After calculating $\widehat{NAPS_\iota}$, the second stage of the instrumental variable methodology is to regress performance variables on the instrumental variable and the shortened matrix of background characteristics. The coefficient on the propensity score, $\widehat{NAPS_\iota}$, is a measure of the impact of treatment on the dependent performance variable. One note to make about using the instrumental variable method is that standard errors are biased and appear smaller than they are. This is a result of using the propensity score as a variable, when it is, in fact, an estimation itself. The propensity score has its own standard error associated with it, so using it as a variable in the second stage regression compounds the errors, but we only observed the measured error from the second stage.

[2] Because USNA is committed to ensuring each graduating class maintains gender, racial, and geographic diversity, these are valid instruments. In order to ensure that the graduating class, after expected attrition, is made up of a diverse group of students from all 50 states and U.S. territories, the admissions process motivates treatment.

IV. Data

The data cover USNA students from 1985 through 2011, a total of 29,939 graduates. The full sample includes 22,743 midshipmen who entered directly from high school, 4796 who went through NAPS, 1929 who went to a Foundation school, and 1039 who attended another college before USNA.[3]

Along with distinguishing between midshipmen who enter the Academy directly and those who first attend NAPS or a Foundation school, the data contain a rich assortment of student characteristics. In terms of background information, the data identify each individual's age, race, gender, SAT scores, high school name, high school location, and high school rank. In terms of potential educational outcome measures, the data include each individual's grades for all courses taken, name of declared major, aptitude grades, and academic, military and overall orders of merit.[4] Summary statistics including a breakdown of treated and control groups by type of precollege program are displayed in Table 1a. and Table 1b.

[3] For information on institutions specifics of NAPS, and Foundation schools, see Appendix 1.
[4] For information on all performance and background variables, see Appendix 3.

Table 1a. Summary Statistics for Key Background Variables

Background Traits	Direct Average	NAPS Average	College Average	Foundation Average
Verbal SAT	651.6	585.3	633.604	636.475
	(63.5)	(65.9)	(67.178)	(58.846)
Math SAT	673.6	603.4	657.012	655.35
	(57.6)	(58.9)	(59.297)	(52.332)
High School Rank (percent)	0.91	0.74	0.836	0.803
	(0.11)	(0.18)	(0.145)	(0.144)
Age on IDay	18.4	19.8	19.677	19.303
	(0.689)	(1.04)	0.892	(0.472)
Central	0.178	0.135	0.154	0.112
	(0.383)	(0.342)	(0.361)	(0.316)
Northern	0.279	0.306	0.278	0.421
	(0.449)	(0.461)	(0.448)	(0.494)
Pacific	0.155	0.174	0.152	0.206
	(0.362)	(0.379)	(0.359)	(0.405)
Western	0.134	0.132	0.152	0.079
	(0.34)	(0.339)	(0.359)	(0.27)
Southern	0.235	0.237	0.25	0.174
	(0.424)	(0.425)	(0.433)	(0.379)
Varsity Athlete (indicator)	0.37	0.47	0.29	0.415
	(0.48)	(0.5)	(0.454)	(0.493)
Military Father	0.427	0.427	0.356	0.452
	(0.495)	(0.495)	(0.479)	(0.498)
Military Mother	0.031	0.051	0.033	0.028
	(0.174)	(0.22)	(0.178)	(0.165)
High School Quality Measure	597.82	467.719	525.808	498.802
	(110.045)	(104.371)	(119.588)	(98.214)
Number of Observations	20,629	4796	1039	1929

Standard deviations appear below observations in parenthesis
See Appendix 3 for variable descriptions.

Table 1a. includes information on student background profiles. These background characteristics should be interpreted as follows: *Verbal SAT* is a record of the student's highest reported verbal SAT score. *Math SAT* is a record of the student's highest reported math SAT score. *High School Rank (percent)* is a percentile rank of each student within his or her respective high school class. A rank of 1.00 means the student was ranked top in his/her class. A rank of .50 means the student was ranked in the very middle of his/her graduating high school class: for example, 45th out of 90 students. The variable *Age on IDay* gives an accurate measure of a student's age on the day he/she reported with the rest of his/her class to Induction Day at USNA.[5] The next five variables are geographic indicator variables indicating 1 for the region in which a student's home of record is located. The geographic regions are broken down as follows: "*Central*" states include Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, and Wisconsin. "*Northern*" states include Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. "*Pacific*" states include Alaska, Arizona, California, Hawaii, Nevada, Oregon, Utah, and Washington. "*Southern*" states include Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia. Finally, "*Western*" states include Colorado, Idaho, Kansas, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas, and Wyoming. In model specifications, *Central* is the excluded variable and when using all state dummies *Alaska* is the excluded variable. *Varsity Athlete (indicator)* is an indicator variable for whether an individual participated on a Varsity Sports team for any of the semesters while at USNA. *Military Mother* and *Military Father* are both indicator variables for whether the individual's parents ever served in any of the armed forces. *High School Quality Measure* is a variable that analyzes the academic quality of individual high schools on the same scale as the SAT: from 200 to 800.

[5] Induction Day refers to the report day for incoming freshman about to begin their indoctrination summer at USNA.

Table 1b. Summary Statistics for Key Variables – USNA Performance

USNA Performance	Direct Average	NAPS Average	College Average	Foundation Average
Graduated (indicator)	0.79	0.77	0.815	0.848
	(0.4)	(0.42)	(0.388)	(0.36)
Start Group 1	0.355	0.255	0.357	0.278
	(0.478)	(0.436)	(0.479)	(0.448)
Start Group 2	0.211	0.261	0.194	0.192
	(0.408)	(0.439)	(0.396)	(0.394)
Start Group 3	0.274	0.359	0.341	0.42
	(0.446)	(0.48)	(0.474)	(0.494)
End Group 1	0.313	0.189	0.291	0.235
	(0.464)	(0.391)	(0.454)	(0.424)
End Group 2	0.2	0.228	0.186	0.181
	(0.4)	(0.42)	(0.389)	(0.385)
End Group 3	0.281	0.36	0.339	0.431
	(0.449)	(0.48)	(0.474)	(0.495)
Major Switch	0.898	0.961	0.909	0.94
	(0.302)	(0.193)	(0.288)	(0.237)
AC grades1	2.703	2.34	2.847	2.544
	(0.677)	(0.583)	(0.68)	(0.595)
AC grades1	2.728	2.28	2.777	2.491
	(0.644)	(0.567)	(0.654)	(0.58)
AC grades1	2.87	2.34	2.916	2.585
	(0.685)	(0.627)	(0.704)	(0.624)
AC grades1	2.902	2.39	2.964	2.651
	(0.661)	(0.586)	(0.614)	(0.592)
Normalized OOM	0.454	0.696	0.472	0.568
	(0.281)	(0.238)	(0.276)	(0.26)
Normalized AOM	0.429	0.68	0.478	0.563
	(0.284)	(0.262)	(0.279)	(0.276)
Normalized MOM	0.445	0.628	0.458	0.492
	(0.29)	(0.277)	(0.281)	(0.287)
Academic Average	2.941	2.513	2.984	2.703
	(0.505)	(0.409)	(0.499)	(0.434)
Military Average	3.318	3.19	3.41	3.314
	(0.323)	(0.309)	(0.316)	(0.32)
Professional Average	3.322	2.946	3.271	3.144
	(0.393)	(0.367)	(0.401)	(0.359)
Youngsterdrop	0.111	0.151	0.1	0.086
	(0.314)	(0.358)	(0.3)	(0.28)
Plebedrop	0.053	0.059	0.054	0.044
	(0.223)	(0.235)	(0.225)	(0.206)
Number of Observations	20,629	4796	1039	1929

Standard deviations appear below observations in parenthesis
See Appendix 3 for variable descriptions.

Table 1b. includes information on performance variables for student assessment while at USNA. Outcome variables should be interpreted as follows: *graduation* refers to the graduation rate of each cohort. *Start Group 1* is a binary variable giving a value of one for a student who originally elects to pursue a degree as a group one major. Group 1 majors include all of USNA's engineering majors. It follows that *Start Group 2* and *Start Group 3* are binary variables given a value of one for students who begin their studies at USNA as Group 2 or Group 3 majors. Group 2 majors are USNA's non-engineering but math and science majors. Group 3 majors are USNA's humanities and social science majors. For a full list of USNA's majors by group see Appendix 2. Conversely, *End Group 1, End Group 2* and *End Group 3* are all binary variables which indicate the major group in which students conclude their time at USNA or from which major group they graduate. The variable *Majorswap* is a binary variable that indicates the likelihood of a member of that cohort switching majors while a student at USNA.

The next four variables are measures of academic course GPAs during the first four semesters at USNA. Academic grades include all academic coursework but exclude military and professional course grades.

OOM refers to the student's normalized overall order of merit at graduation. Overall order of merit combines measures including academic and professional course grades, military performance, conduct, physical education, and athletic performance. *AOM* refers to the normalized academic order of merit at graduation for each cohort. Academic order of merit is a measure of academic performance during eight semesters of coursework. *MOM* refers to students' normalized military order of merit at graduation. Military order of merit is a measure of military performance that includes students' military grade, conduct grade, professional course grades, athletic performance, and physical education grade. In the interpretation of these variables, it is crucial to recognize that these are not measures of military, academic, or overall performance in a pure sense. Rather, they describe relative performance compared to the performance of all other students in a graduating class. In the results section, a higher or lower *MOM*, *AOM*, or *OOM* signifies a student performing better or worse relative to his/her classmates.

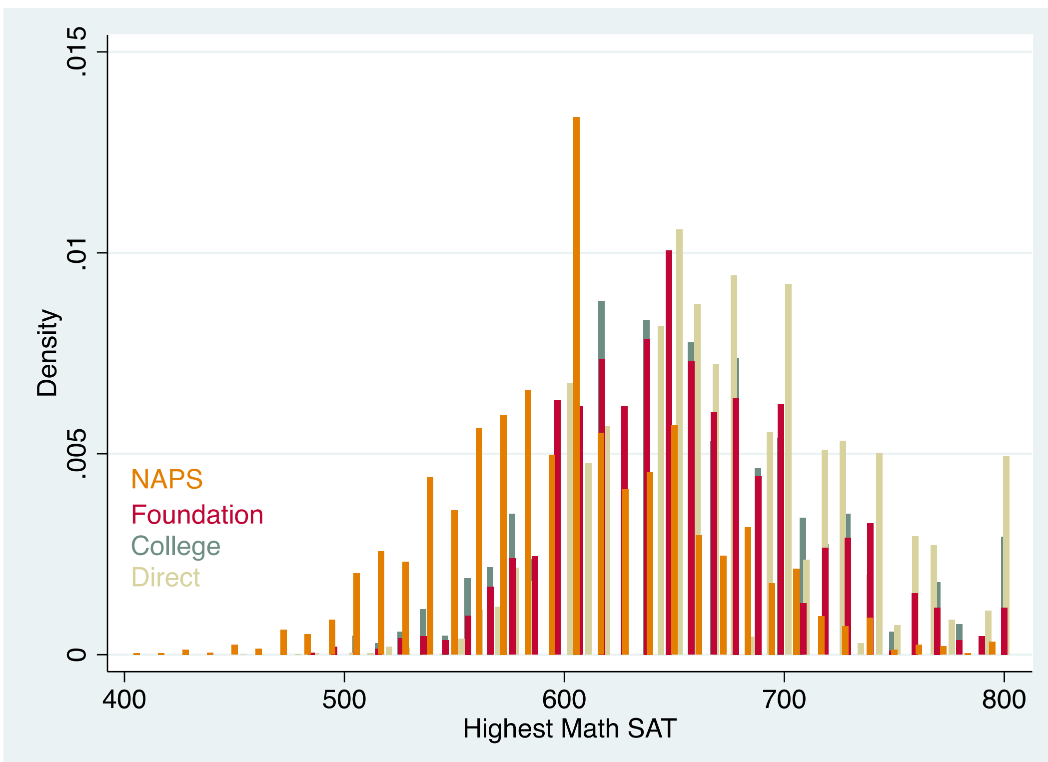
The next three variables are measures of academic, military, and professional course GPA averages over four years of study. The first, titled *Academic Average,* refers to the average course GPA in all academic courses for the first four semesters excluding professional and military coursework. The second, titled *Military Average*, refers to military performance measures including physical education grades, conduct grades, professional course grades, and military performance grades. The supervising officer assigns the military performance grade. Finally, the variable *Professional Average* is a measure of an individual's average professional grades over four years at USNA.

The variables *Youngsterdrop* and *Plebedrop* are binary variables that indicate the likelihood of a member of that cohort leaving USNA as either a sophomore or freshman.

Consider the summary statistics above. First note that average differences between treated and non-treated groups for variables such as SAT scores and high school rank are quite large. For example, the average math SAT scores for incoming high school students is 70 points higher than those of students who first attend NAPS. Table 1a. and Table 1b. display key differences in both background characteristics and performance metrics between treated and control groups before matching. Simply looking at the averages for background characteristics, we find that there appears to be a dramatic difference in the magnitude and spread of the high school percentile variable and the high school quality rank variable. We also see a significant difference in the means of performance variables like academic course grades. The difference in means but large overlap in terms of standard deviation is what motivates an examination of similar characteristics when examining only a matched cohort.

It is useful to see graphically how the overlap in data motivates the use of propensity score matching. Figure 1 illustrates the overlap in Math SAT scores among the four feeder sources: NAPS, Foundation, college, and direct entry.

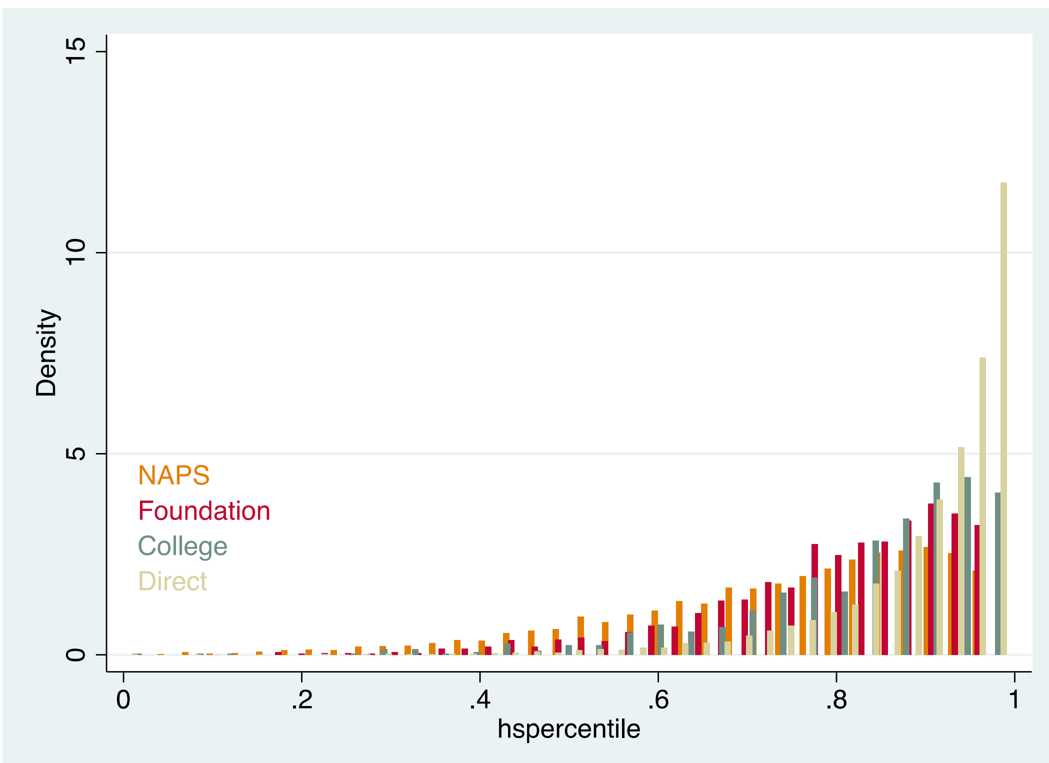
Figure 1 – Math SAT Score Density by Feeder Source



This graph illustrates the dramatic amount of overlap in math SAT scores between the feeder program individuals and the direct entry individuals. This result is also true for verbal SAT scores and for high school percentile rank (see Figure 2). The significant overlap is perfect for matching propensity scores between groups in order to analyze unbiased treatment effect.

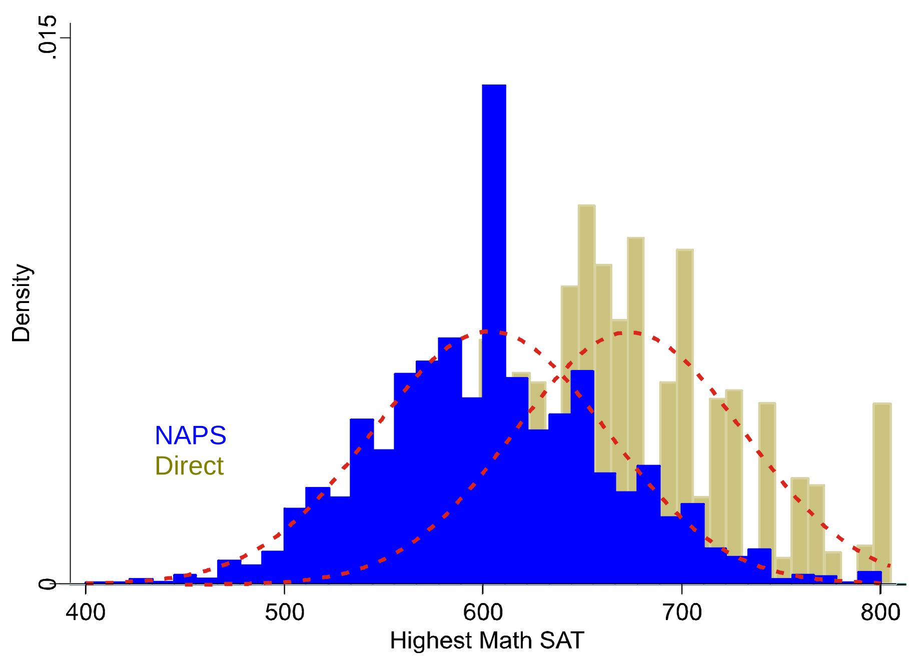
After individuals from the treatment and control groups are matched based on propensity scores, we estimated the treatment effect by comparing the mean outcome values of the matched cohort for the treatment group and control group.

Figure 2 – High School Percentile Density by Feeder Source



Again if we consider Table 1a. and Table 1b., while the mean values for characteristics appear to demonstrate clear demarcation lines between treated and control groups, standard deviations among the variables are large, suggesting a great deal of *overlap* in student characteristics between the two groups. Consider, for example, the distributions of math SAT scores between treated (NAPS) and non-treated, displayed in Figure 3. The large section of overlapping scores between groups informs our decision about which analysis techniques to use. First, a discontinuity design based strictly on SAT scores would not work. Second, the admissions office was very likely using much more information than just SAT scores in determining selection. For our purposes, the overlap that exists between groups in many student characteristics allowed for the use of propensity score matching to properly estimate treatment effects.

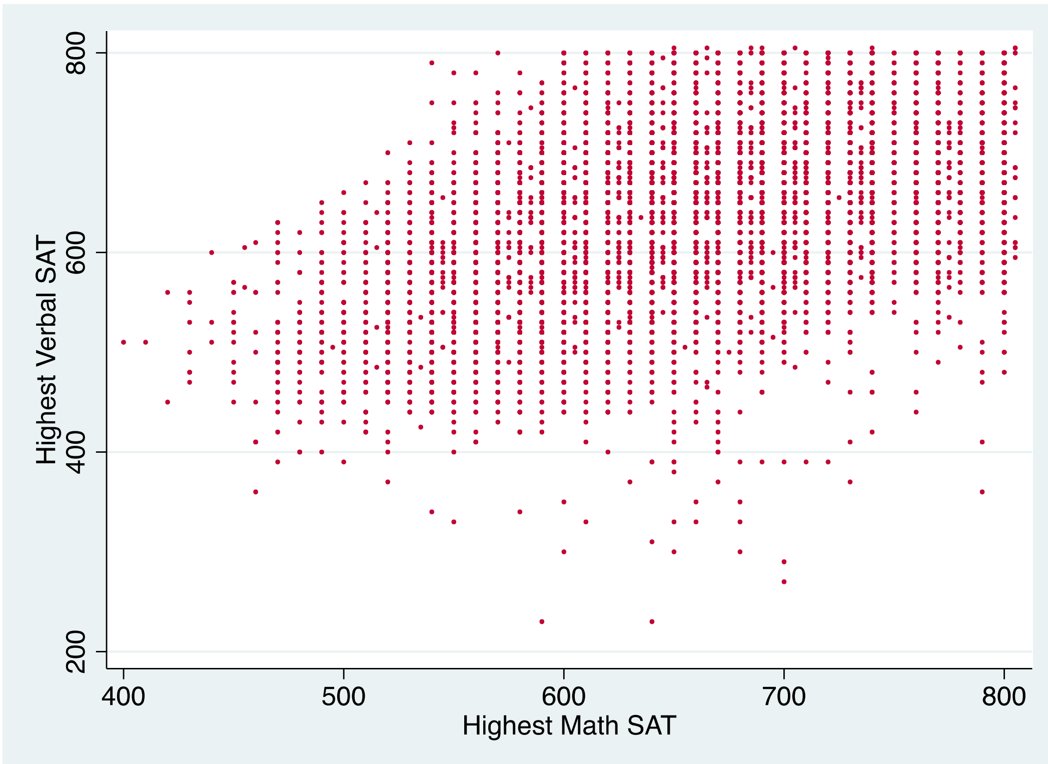
Figure 3 – Distributions of Math SAT Scores for NAPS (Blue) and Direct (Gold) Students



Figures 1, 2, and 3 illustrate the overlap in background characteristics across the different feeder sources. As discussed earlier, one of the assumptions required for use of propensity score matching is the Common Support Condition (CSC), which states the probability of being assigned treatment falls between 0% and 100% for any background characteristic X. The overlap in background characteristics illustrated by Figures 1, 2 and 3 legitimizes the use of propensity score matching as a method by which to mitigate selection bias and assess the effects of treatment.

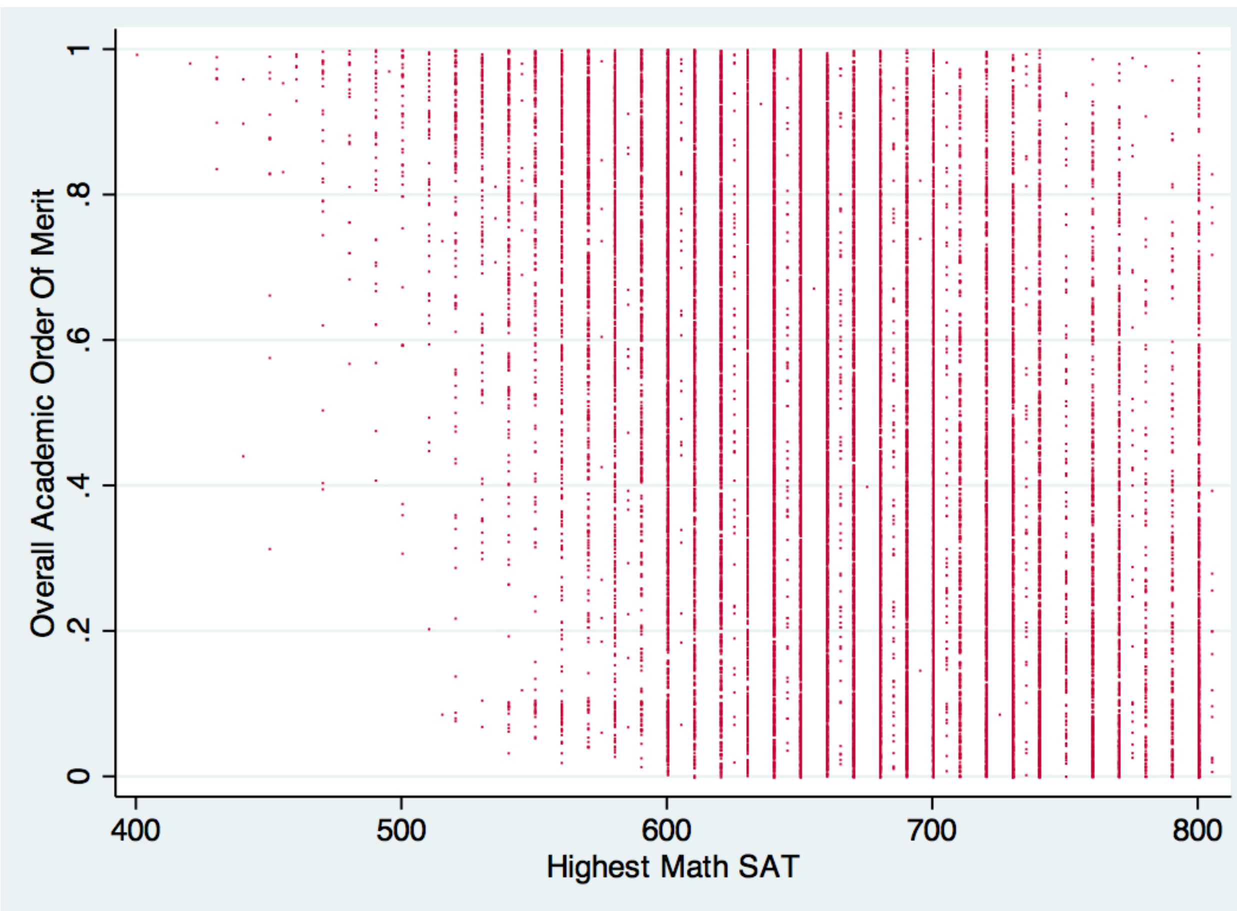
Examining scatterplots of correlations not only reinforces the hypothesis that there is no definite split between background characteristics of treated and control individuals but suggests that individual observations fall all over the map, with only cloudy correlation. For example, Figure 4 depicts the cloud of observations that represent the positive correlation between Math SATs and Verbal SATs. While there is an obvious positive trend depicting the relationship between SAT scores, there is no clear line of demarcation at which the admissions board would be able to make a cut-off value for direct admission. The data are spread out into one large cloud of observations with a large variance. This is evidence that the CIA assumption previously discussed is fulfilled by the dataset and helps inform variable selection.

Figure 4 – Correlation Between Math SAT scores and Verbal SAT scores



There is massive variation in the data, which both eliminates the possibility of having an unbiased comparison of two groups based on the mean, while simultaneously setting up an ideal scenario for propensity score matching. This variance is mirrored in correlation comparisons of data over time. Figure 5 illustrates a similar cloud of varied observations between overall academic order of merit and Math SAT scores. Figures 4 and 5 illustrate the heterogeneity in background characteristics. For example, Figure 5 illustrates the high density of individuals with SAT scores in the range between 500 and 800 whose overall academic orders of merit (*AOM*) vary from being in the top 10% to the bottom 10% by class. Heterogeneity in background characteristics emphasizes the importance of selecting a model specification that includes multiple variables and background characteristics. Even though we might assume that if an individual got a very high score on the math SATs, he/she would also get a very high score on the verbal SATs, the figures above would disprove our assumption. Therefore, it is key to include both SAT variables in our analysis. The multidimensionality of background characteristics is a crucial aspect of the data to consider when generating propensity scores. Finally, figures 4 and 5 demonstrate the fact that multi-collinearity should not create much of an issue when undertaking OLS, logistic, or IV regressions. While background characteristics like Math and Verbal SATs are correlated with each other as well as with USNA performance metrics, figures 4 and 5 illustrate the accompanying heterogeneity.

Figure 5 – Correlation Between Math SAT scores and Overall Academic Order of Merit



V. Model Estimation

This paper utilizes three methods to assess returns from pre-college programs: ordinary least squares (OLS) regression, propensity score matching (PSM), and instrumental variable (IV) analysis. OLS regression is designed to assess a linear relationship between the treated and control groups. Although it is important to include in this paper, OLS gives results that are potentially biased because it relies on the average between the entire treated and control groups. By using PSM analysis, we can specify the cohort of interest to include only those individuals with similar background characteristics who either attended or did not attend a pre-college program. In this way, PSM eliminates students who, based on their background characteristics, either would never or would always have been assigned the treatment of a pre-college program. This allows for a closer examination of the direct effects of the treatment on treated students. Finally, the propensity score generated in PSM is a useful tool for the third method of treatment analysis. The propensity score allows for regression analysis using the propensity score as an instrument. This IV analysis is a method of regression analysis that provides a more specified estimate of the impact of pre-college programs.

In deciding which variables to use as the determinants in analyzing returns to pre-college education, we considered information about the USNA admissions process to inform our analysis. The USNA admissions board relies on a measure called the "candidate multiple" (CM) to weigh a student's background characteristics and potential for success as a Midshipman. The criteria that fall into the CM are broken down by percent. Rank in high school class is given the highest weight, at 27% of the CM. Because USNA is an engineering school from which all students graduate with a Bachelor of Science, admissions decisions weigh heavily in favor of an individual's propensity to succeed in high level math and science courses. For this reason, math

SAT scores make up 24% of the CM, but verbal SAT scores make up only 12%. Technical interest, or a student's expressed desire to study a technical subject, accounts for 14% of the CM. Official high school recommendations are weighted 11%. High school extra-curricular activities are weighted at 8%, and the final 4% of CM weight is assigned according to whether a student expresses an interest in a career in the military (Fitzpatrick, 2001). This information on the candidate multiple suggests which variables should be included in OLS, PSM, and IV analysis. In addition, the dataset used in this analysis includes the specific key characteristics that are used by the admissions board to determine the CM.

A. Ordinary Least Squares and Logistic Regression

Equations for the OLS and logistic regression models are based on examining the effect of the binary treatment variable (either NAPS, Foundation, or College) on a range of performance variables. For OLS and logistic regressions, equations take the following forms:

$$Y_{it} = \alpha + \beta * (NAPS_{it}) + \delta_1 * \mathrm{MathSAT_{it}} + \delta_2 * \mathrm{VerbalSAT_{it}} + \delta_3 * \mathrm{HSRank_{it}} + \delta_4 * \mathrm{HSQuality_{it}} + \varepsilon_{it} \tag{7}$$

Where Y_{it} is AOM_{it}, $GradStat_{it}$, $AcGrade1_{it}$, $AcGrade2_{it}$, $AcGrade3_{it,}$ and $AcGrade4_{it}$. The OLS equations include a binary variable that indicates participation in NAPS, controls for Math and Verbal SAT scores, high school rank, and high school quality. The coefficient β is the measure for the impact of NAPS on the dependent performance variable. The performance variables which we examined using OLS are academic order of merit, graduation rate, and academic grades during the student's first four semesters as a midshipman.

B. Propensity Score Matching

In propensity score matching, the first stage equation is a regression similar to those above. In the second stage, we employed a matching algorithm to match individuals from the treatment and control groups based on their calculated propensity scores. Propensity score matching allows us to examine more closely the direct impact of NAPS on the cohort of students who are treated. Unlike OLS, PSM examines only those students who fall into the matched cohort of treated and untreated students. Similar to the regressions above, the USNA admissions policies suggest the use of four key background variables in calculating the propensity score. In the context of the Conditional Independence Assumption (CIA), the information on the candidate multiple determines the correct specification for background characteristics that we used in order to mitigate selection bias. Admissions puts heavy weight on high school rank and SAT scores. Therefore, these are the characteristics that most directly affect selection into the treatment group, or assignment to a pre-college program, and therefore are included in our first stage propensity score matching model.

Matching criteria for the first stage propensity score must impact both the selection for treatment and the final outcome variable (Caliendo and Kopeinig, 2008). However, variables must not be affected by treatment and therefore must be fixed over time or measured before treatment. Over-specifying the first stage model can exacerbate what is referred to as the common support problem: where there is no overlap in background characteristics for treated individuals in the un-treated sample, those treated individuals are dropped from analysis. This is a danger of propensity score matching; there is the potential for eliminating a significant number of treated observations, leading to incomplete results. At the same time, isolation of a specific cohort of the data is part of what informs returns to treatment and eliminates selection bias

(Bryson, Dorsett and Purdon, 2002). Specifically, this paper compares the academic performance of students with background characteristics that are comparable to those of students who actually attended preparatory school, rather than to all students admitted to USNA. Finally, there is danger in over-specifying the model and inflating variance (Lechner and Smith, 2002; Bryson, Dorsett, and Purdon, 2002). Others suggest that all variables related to the outcome that are proper covariates should be included in any first stage specification (Rubin and Thomas, 1996).

In this study, we chose to run a very inclusive specification and a second specification that only included a few choice matching criteria. This provides a robustness test as well as ensuring that the first stage is an accurate model on which to predict the second stage. For the less inclusive specification, we used Math SATs, Verbal SATs, high school rank, and high school quality in the first stage equations. Equations for the propensity score matching section of this paper take the following forms:

$$Y_{it} = \alpha + \delta_1 * \mathrm{MathSAT_{it}} + \delta_2 * \mathrm{VerbalSAT_{it}} + \delta_3 * \mathrm{HSRank_{it}} + \delta_4 * \mathrm{HSQuality_{it}} + \varepsilon_{it} \qquad (8)$$

Where $Y_{it}$is $NAPS_{it}$, $Foundation_{it}$, or $College_{it}$. In this case, the propensity score is the estimated value of $\widehat{NAPS}_{it}$, $\widehat{Foundation}_{it}$, or $\widehat{College}_{it}$. In the second stage, we employed nearest neighbor, caliper, and kernel matching to compare the difference in means of the propensity score of the treatment and control group between matched observations.

C. Instrumental Variable Regression

In the final method of comparing treatment and control groups, we employed the propensity score as an instrumental variable. This comparison again utilizes the propensity score. In order for the propensity score to be a valid instrument, it must include information that is not directly attributable to success at USNA but directly influences admission to NAPS. This is the definition of the exclusion principle. In this case, the inclusion of the state variable makes $\widehat{NAPS}_{it}$ a valid instrument. Due to the mandate that USNA graduate a diverse class of officers who reflect the makeup of the fleet and the commitment of USNA to graduating officers from every state and US territory, USNA admissions wants to ensure that students from all states and territories are successful enough at USNA to graduate. Admissions can directly influence this success by assigning individuals to pre-college programs. Therefore, a student's home state effects his/her likelihood of assignment to pre-college but not directly his/her performance at USNA. The first stage equation for IV analysis is in the following form:

$$NAPS_{it} = \alpha + \delta_1 * \mathrm{MathSAT_{it}} + \delta_2 * \mathrm{VerbalSAT_{it}} + \delta_3 * \mathrm{HSRank_{it}} + \delta_4 * \mathrm{HSQuality_{it}} + \delta_5 * \mathrm{Varsityathlete} + \delta_6 * \mathrm{Gender} + \delta_7 * (Ethnicity\ Dummy) + \delta_8 * (State\ Dummy) + \delta_9 * (Year\ Dummy) + \varepsilon_{it} \tag{9}$$

After calculating $\widehat{NAPS}_{it}$, we insert those values into a regression formula and observe the coefficient on that variable. The second stage models are in the following forms:

$$Y_{it} = \alpha + \psi * (\widehat{NAPS}_{it}) + \delta_1 * \mathrm{MathSAT_{it}} + \delta_2 * \mathrm{VerbalSAT_{it}} + \delta_3 * \mathrm{HSRank_{it}} + \delta_4 * \mathrm{HSQuality_{it}} + \varepsilon_{it} \tag{10}$$

Where $Y_{it}$is AOM_{it}, $GradStat_{it}$, $AcGrade1_{it}$, $AcGrade2_{it}$, $AcGrade3_{it,}$ and $AcGrade4_{it}$. The variable ψ is the coefficient on $\widehat{NAPS}_{it}$, which gives an estimate of the treatment effect on performance variables.

VI. Empirical Results

Across all three methods of analysis, qualitative results remain consistent. Pre-college education programs, both NAPS and Foundation school, are associated with positive returns to college graduation rates. However, students who attend these pre-college programs will perform at a lower level academically overall than similar students who do not attend pre-college programs. Moreover, results indicate positive returns to academic grades during the first semesters at USNA, followed by insignificant results, and in PSM analysis, negative returns to academic grades in the third and fourth semesters at USNA.

A. Ordinary Least Squares and Logistic Regression

Before we undertake any regression methods that expressly address selection bias, it is useful to examine how ordinary least squares and logistic regression describe returns to educational outcomes based on participation in a preparatory school program. Table 2 illustrates the coefficients on the binary independent variable *NAPS* in OLS and logistic regressions with each of the performance variables listed as dependent variables, and *NAPS, HighestMathSAT, HighestVerbalSAT, hspercentile, and HsOfficialStClassRank* as the independent variables.

Coefficients on the *NAPS* variable paint an interesting picture. Positive returns to NAPS participation on academic grades during the first semester (0.1475) diminish in magnitude and eventually become negative by the third semester (-0.0341). Returns on military performance are negative by the second semester. *Academic order of Merit* is negatively impacted by attendance at NAPS, causing a decrease in normalized *AOM* of 2.76%. Logistic results suggest that those who attend NAPS are less likely to drop out controlling for background characteristics. In fact, individuals are 1.3 times more likely to graduate having gone to NAPS.

The OLS and logistic regression methods of analysis are important tools but may fail to explicitly account for the endogeneity associated with this type of treatment study. Since the regression coefficients are based on the entire data sample, a linear relationship between the background variables and the probability of going to NAPS is assumed. Ordinary least squares regression is a method with fundamental pitfalls for this type of analysis, which has both a treatment and control group. Results from a linear regression might be extremely sensitive to the averages of each group. If we consider individuals on the very high end of performance metrics, there will be observations in each group who would inevitably go to NAPS or not go to NAPS based on their background characteristics. For example, an individual with perfect SAT scores ranked first in his/her high school class has a very low probability of being assigned to a remediation program. Conversely, an individual with 400 SATs for math and verbal and a very low high school rank has a very low probability of being admitted directly to USNA. However, when using OLS or logistic regression and trying to create a linear relationship between treatment and control groups, the observations that are unlikely to appear in either group might skew the linearity of the regression and affect the averages. For this reason, OLS and logistic regressions could produce biased estimates of treatment effects. However, it is still useful to analyze the results of OLS and logistic regressions because they provide a baseline result from which we can turn to PSM and IV analysis.

Table 2 - OLS and Logistic Regression - Impact of NAPS on Outcome Variables

	Grad AOM	Graduation Rate (Logistic)	Ac Grade 1	Ac Grade 2	Ac Grade 3	Ac Grade 4
NAPS	0.0276***	1.303***	0.1475***	0.0232**	-0.0341***	-0.078***
	(0.0056)	(0.063)	(0.0119)	(0.0116)	(0.0127)	(0.0127)
Highest Math SAT	-0.0011***	1.004***	0.003***	0.0025***	0.0035***	0.0026***
	(0.00003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Highest Verbal SAT	-0.0007**	0.9996	0.0015***	0.00138***	0.001***	0.0011***
	(0.00003)	(0.0003)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
High School Rank (Percentile)	-0.0171	2.339***	0.0506	0.0578	0.0481	0.0258
	(0.0247)	(0.461)	(0.0513)	(0.0502)	(0.0553)	(0.0555)
High School Quality Measure	-0.0007***	1.0001	0.0014***	0.0014***	0.0012***	0.0012***
	(0.00003)	(0.0002)	(0.00006)	(0.00006)	(0.0001)	(0.0001)
Number of Observations	20529	25988	24046	23518	22811	22255
R-squared	0.2903	Pseudo: 0.0095	0.2434	0.2493	0.246	0.2145
Adjusted R-squared	0.2902	--	0.2432	0.2491	0.2458	0.2143

Standard errors in parentheses.
*** indicates significance at the .01 level. ** indicates significance at .05. * indicated significance at .1.
See Appendix 3 for variable descriptions.
For Graduation Rate, coefficients represent odd ratios.

The coefficients on the independent variables *Math SAT, Verbal SAT, High School rank,* and *High School quality* have the expected signs and magnitudes. Higher Math and Verbal SAT scores are associated with a higher *AOM*, higher likelihood of graduation, and positive returns in all four semesters of academic grades. *High School rank* and *High School quality* are also associated with positive returns to *AOM*, *Graduation Rate*, and all four semesters of academic grades.

B. Propensity Score Matching First Stage – Creating a Matching Specification with Background Estimators

Table 3 – Coefficients for First Stage Matching Using NAPS

	Model 1	Model 2	Model 3
High School Quality Measure	-0.0036***	-0.004***	-0.0035***
	(0.0002)	(0.0003)	(0.0003)
High School Rank (Percent)	-0.886***	-0.358***	-0.374***
	(0.129)	(0.166)	(0.218)
Verbal SAT	-0.006***	-0.005***	-0.005***
	(0.0003)	(0.0003)	(0.0003)
Math SAT	-0.0098***	-0.008***	-0.008***
	(0.0003)	(0.0004)	(.0004)
Age on IDay	--	1.33***	1.35***
		(0.027)	(.0274)
Hispanic	--	0.821***	0.877***
		(0.050)	(.052)
Asian American	--	0.699***	0.71***
		(0.084)	(.086)
African American	--	0.789***	0.791***
		(0.055)	(0.056)
Sex	--	0.232***	0.229***
		(0.0496)	(0.05)
Western	--	--	0.026***
			(.067)
Southern	--	--	0.356***
			(.058)
Pacific	--	--	0.240***
			(.064)
Northern	--	--	0.444***
			(.056)
Number of Observations	21116	21116	21116
R-squared	0.3996	0.6077	0.6152

Dependent variable is an indicator variable for whether a student attended NAPS.
Standard errors in parentheses.
*** indicates significance at the .01 level. ** indicates significance at .05. * indicated significance at .1.
See Appendix 3 for variable descriptions.

Table 3 illustrates the first stage of the two-stage PSM analysis. In the first stage, we generated the propensity score that is then algorithmically matched in the second stage. For each specification, all coefficients are statistically significant. The first stage coefficients suggest that gender, race, and age are much more influential in determining whether an individual will attend NAPS than his/her SAT scores are. The direction of coefficients suggests that being female, a minority, or a year older increases the probability of the student being sent to NAPS. This is a scenario in which the specificity of the model is causing certain coefficients to absorb the impacts of variables not included in the model. In this case, the model is over-specified to include variables that should not have an impact on the graduation rate. Gender and ethnicity should theoretically have little impact on outcome measures like academic grades or graduation rates. However, it is still useful to analyze the effect of these variables and consider what other factors they are absorbing, particularly family income and parents' education level. These variables will show up again as part of the instrumental variable estimation section to satisfy the exclusion restriction.

The first model is the primary model of interest for this study. The admissions board weighs SAT scores and high school class rank the heaviest when assigning a CM score to each candidate. Therefore, it follows that our specification for assessing the probability of being assigned to NAPS, offered a place at a Foundation school, or just rejected from USNA entirely should depend upon those same key metrics. The coefficient magnitudes and directions follow from using logic about the definition of pre-college education and the purpose of NAPS. SAT scores are accompanied by small negative changes in the propensity to go to NAPS. High school rank absorbs most of the model effects, suggesting that being ranked highly in one's graduating high school class is the strongest contributing factor as to whether a student is offered an appointment to NAPS. Finally, the relative rank of the high school one attends contributes marginally to attendance at NAPS. Those who attend more highly ranked high schools will be marginally less likely to be sent to NAPS. Of note, the low coefficient of determination, R squared, suggests that 39.9% of the variation on the left side of the regression equation is explained by variation on the right side of the equation.

Model 2 includes the background characteristics of sex, ethnicity, age, SAT scores, high school class rank, and high school quality measure. This is a very inclusive specification and runs the risk of generating results with high variance. It also includes variables that theoretically have no impact on the outcome measures. Gender and ethnicity theoretically have little impact on outcome measures like academic grades or graduation rates. However, for robustness purposes it is crucial to include this specification as a first-stage model.

When we eliminate gender, race, and age from the model, the magnitude of coefficients is mostly absorbed by the high school rank variable. This follows the conclusion that the standard errors are serving as proxies for variables for which we have no data. In a similar way, a measure for age is a double-edged sword when included in the first stage specification. NAPS (and Foundation school) students tend to be older than the typical midshipman who enters the Academy straight from high school. There is the possibility that any positive effects in terms of academic performance are simply a result of the fact that they are one year older and are therefore more likely to succeed in college simply because they have greater maturity. However, the data includes significant overlap in age between treated and untreated populations, so a simple control should prevent any systematic error. At the same time, the Age on Induction Day variable strongly predicts whether a student attended a pre-college program, introducing bias that may skew results. Despite the inclusion of seemingly irrelevant variables, the first model has an R squared value of .6077. In other words, 60% of the variation in the dependent variable is explained by variation in the independent variables.

Finally, the third model includes a geographic region proxy variable based on the region in which the individual went to high school. The specification is relative to the *central* variable. Again, this variable should not have a direct impact on graduation rates, but education by state and by region of the United States is very different. However, even with the inclusion of these new variables, the R squared value stays relatively constant at .6152. The second-stage results for each specification are highlighted in the next section of this paper.

C. Propensity Score Matching Second Stage – Matched Results

The results of this study are broken down into three major sections: results for NAPS students, results for Foundation school students, and results for students who attended a four-year college before matriculating at USNA.

The following tables document the cohort differences in a number of outcome variables between NAPS students and direct entry students, Foundation students and direct entry students, and prior college students and direct entry students.

For robustness purposes, we include five matching methods for each matched outcome variable. NN(1), NN(5), and NN(20) refer to nearest neighbor matching results using the closest nearest neighbor, the five nearest neighbors, and the 20 nearest neighbors. Caliper refers to caliper matching with a radius of .001 units. Kernel refers to kernel matching using the normal Gaussian method.

i. NAPS

Table 4a illustrates second stage matched results comparing NAPS and direct entry students. If we consider the aim of the NAPS program is to prepare students for a rigorous four years of academic work at USNA, the key performance metrics by which we measure success are graduation rates and academic performance metrics. As documented in Table 4a, the cohort of NAPS students graduate at a rate nearly 10% higher than those who do not attend a precollege program controlling for background characteristics. Even the most modest estimate of positive returns to graduation rates, using only the single nearest neighbor method, generated an estimate that those who attend the NAPS program graduate at a rate 7.3% higher than those students of similar backgrounds who do not attend the NAPS program. This is a significant result as it suggests positive returns to the NAPS program for a key performance metric. Graduating as many students as possible is not only in the interest of the U.S. Navy but is an expressed goal of national education policy makers. As was discussed in the introduction to this paper, degree attainment is a determinant of future wage earnings.

Although returns to graduation rates are positive, returns to class rank are significantly negative. Matched results suggest that overall order of merit is negatively impacted by attendance at the NAPS program. Matching based on the 20 nearest neighbors suggests that overall order of merit is decreased by 4.2% for NAPS attendees compared with direct entrants. Academic order of merit is similarly affected by attendance at NAPS. Academic order of merit is decreased by over 6% based on all four matching methods. Military order of merit is even more negatively impacted, since it decreased by 6.5% using the most modest estimate. One conclusion from these results is that the NAPS program helps with retention rates of marginal students while failing to help them improve academically. There are a few possible explanations for these results. It is possible that pre-college programs like NAPS and Foundation schools imbue students with skills that may not help them earn higher grades but help them persevere through a difficult college experience. This could be a result of students being one year older and more mature than they were coming directly out of high school. Another explanation for graduation rates is that NAPS students feel added pressure to finish their degree program at USNA because they have already committed an extra year to the program. Meanwhile, the academic performance generalizations could be a result of NAPS students seeing similar course material repeatedly. For example, a NAPS student might take calculus during their senior year of high school, again at NAPS, and then again at USNA during their first semester as a Midshipman. Familiarity with the material could explain higher performance during the first semester. Then, once students begin taking unfamiliar courses during the second semester, their higher performance diminishes.

An unmatched academic grades comparison indicates significant negative returns to the NAPS program. However, matched results show a trend of positive returns to the NAPS program in the first semester of academic coursework followed by diminishing returns and eventually negative returns in the third and fourth semesters. The magnitude of academic grade improvement in the first semester is only about half as great as the magnitude of improvement for NAPS students in their STEM course grades in the first semester. For example, the most modest estimate of academic grade improvement in the first semester for NAPS students is .152, using only the single nearest neighbor as a match. By the second semester returns are insignificant. In the third semester returns are significantly negative, but results should be interpreted cautiously due to increased heterogeneity. The most modest estimate using the 20 nearest neighbors is still -.095. Returns are also negative in the fourth semester, with the most modest estimate still at a value of -.076. When we analyze academic grades and reasons behind the positive returns in the first semester followed by diminishing and then eventually negative returns, it is crucial to note that major selection occurs at the end of the second semester of an individual’s freshman year. This means that third semester academic course grades include noise associated with each individual picking his/her own major and beginning his/her major courses. These results may also reflect the fact that students take courses in pre-college programs that they then repeat during their first year at college. However, after repeating a course they have already taken, these students find themselves unprepared for new coursework. This could signify superficial positive returns to academic grades in the first semester of college, since students are simply repeating what they have already been exposed to, not demonstrating a higher performance level than the matched cohort of students who did not attend pre-college.

USNA has a mandate to graduate at least 70% of each class with a STEM degree. This puts a large amount of emphasis on academic performance in the STEM fields. Unmatched results comparing grades in STEM courses for the NAPS cohort and the direct entry cohort suggests negative returns during the first four semesters. However, the matched results tell a different story. The matched results indicate that there are positive returns to STEM course grades in the first semester at USNA. These positive returns range from .296 to .316 on a 4.0 scale. The magnitude of the matched results indicates that those who attended the NAPS program have positive returns in their first semester STEM grades on the order of 7.5% higher grades. During the second semester these positive returns diminish to insignificance. By the first semester of the sophomore year at USNA the returns to STEM grades are significantly negative. Between the third and fourth semester returns are between -.113 and -.06. While the magnitude of these negative returns is smaller, they are still statistically significant. These matched results suggest that the NAPS program is not helping students improve their grades in technical courses after the first semester, and due to some aspect of the NAPS cohort, the group performs significantly worse in STEM course grades than individuals with similar backgrounds who entered USNA directly from high school.

Explanations for these negative returns include the possibility of peer effects having an impact on the academic performance of NAPS students. Another explanation is that by the third semester of academic coursework, variance and noise have increased to the point that a realistic comparison of science, technology, engineering, and math GPAs is unreasonable. During the first year at USNA, the vast majority of students are immersed in the same coursework at the same time. This allows for a fairly consistent comparison across all students for the first year. However, by the third semester, students have begun their major coursework and have more freedom in their schedules to take on other classes and participate in other activities. This adds significant noise to the comparison after year one.

Returns to military coursework mirror the same trend. Returns to the NAPS program in the first semester are significantly positive, with a minimum magnitude of .099 in the positive direction using caliper matching. Returns diminish to insignificance, but by the third semester, the beginning of an individual's sophomore year, returns are significantly negative with a minimum magnitude of -.059. By the fourth semester returns on military performance are again insignificant.

These results are somewhat troubling considering the fact that NAPS participants have been immersed in a military style preparatory school program for a year before entering USNA. They have already undergone one military indoctrination summer and have been active duty members of the U.S. Navy for an entire year before they begin courses at USNA. The expectation is that NAPS participants are better prepared for the military aspects of USNA than the comparable cohort of direct applicants. However, results suggest that the positive returns of a year of military experience are of small magnitude and are fleeting, or that non-NAPS students catch up and pass their NAPS peers in terms of military performance.

Effects on major course grades are insignificant in a student's sophomore year. However, during a student's junior year there are negative returns on major's course grades from having attended the NAPS program. Even according to the most modest estimate using the 20 nearest neighbors, majors grades are decreased by -.087 in the first semester of the student's junior year, and -.087 in the second semester, junior year.

The final significant outcome measures include that of an individual's propensity to leave USNA before graduation. *Youngsterdrop* and *plebedrop* both indicate that the NAPS program has a positive impact on retention during sophomore year. Freshman and sophomore retention is of particular interest for USNA because those individuals who are unwilling to commit themselves to service after graduation will typically leave the Academy of their own accord during their first two years before they incur debt. According to matched results, NAPS participants are between 4.1% and 6.1% less likely to drop out during their youngster year and between 4.3% and 6.3% less likely to drop out during their freshman year.

The End Group 1 and *End Group 3* variables are insignificant across the board. However, the *End Group 2* variable suggests positive returns to ending in a group 2 major if a student attends NAPS. This significant result may reflect the fact that group 2 includes the general science major. NAPS students are between 5.6% and 6.8% more likely to be a group 2 major. Interestingly, of the 1049 group 2 majors who graduated from USNA after attending NAPS, 276 of them majored in General Science (26%). However, out of the 4389 group 2 majors who graduated from USNA without attending a pre-college program, only 363 majored in General Science (8%). This leads to the conclusion that while NAPS students are more likely to become group 2 majors, they are overwhelming graduating with a General Science degree.

Finally, the rate at which students swap majors is a representation of academic preparedness for a demanding level of coursework. Matched results indicate that the cohort of NAPS students switch majors between 9.8 and 11.7 percentage points more often that direct entry students.

Propensity score matching leads to the conclusion that there are significant positive returns to investment in the NAPS program. Students are better prepared for (or at least more familiar with) their first semester coursework. There are also significant positive returns to graduation rates and retention. From a macroeconomic viewpoint, this conclusion is crucial for the implications of the NAPS program. If a year of pre-college is improving graduation rates by 10%, many marginal students could benefit from similar treatment. As long as students graduate and receive a degree, they will be part of a different workforce with more resources and available jobs, regardless of whether their pre-college prepared them academically.

Table 4a – Second Stage Matched Results Comparing NAPS and Direct Entry Students

	Unmatched	NN(1)	NN(5)	NN(20)	Caliper	Kernel
Graduation	-0.008***	0.073***	0.101***	0.092***	0.086***	0.083***
	(0.009)	(0.022)	(0.017)	(0.016)	(0.017)	(0.014)
OOM	0.278***	0.048***	0.04***	0.042***	0.053***	0.062***
	(0.007)	(0.014)	(0.012)	(0.011)	(0.011)	(0.01)
AOM	0.295***	0.068***	0.065***	0.068***	0.062***	0.086***
	(0.007)	(0.015)	(0.012)	(0.011)	(0.012)	(0.011)
MOM	0.231***	0.072***	0.065***	0.07***	0.077***	0.08***
	(0.007)	(0.016)	(0.013)	(0.012)	(0.012)	(0.011)
AC grades1	-0.411***	0.152***	0.162***	0.164***	0.16***	0.112***
	(0.014)	(0.031)	(0.025)	(0.024)	(0.024)	(0.023)
AC grades2	-0.516***	-0.002	-0.008	-0.023	-0.014	-0.066**
	(0.014)	(0.028)	(0.024)	(0.022)	(0.022)	(0.022)
AC grades3	-0.567***	-0.146***	-0.098***	-0.095***	-0.119***	-0.132***
	(0.015)	(0.033)	(0.027)	(0.025)	(0.026)	(0.024)
AC grades4	-0.542***	-0.111***	-0.076***	-0.093***	-0.106***	-0.13***
	(0.014)	(0.031)	(0.026)	(0.024)	(0.024)	(0.023)
MIL grades1	-0.033***	0.113***	0.113***	0.121***	0.099***	0.104***
	(0.01)	(0.026)	(0.02)	(0.019)	(0.02)	(0.017)
MIL grades2	-0.219***	-0.025	0.009	-0.006	-0.028	-0.022
	(0.01)	(0.026)	(0.021)	(0.019)	(0.02)	(0.017)
MIL grades3	-0.228***	-0.068**	-0.089***	-0.087***	-0.059**	-0.098***
	(0.011)	(0.029)	(0.024)	(0.022)	(0.023)	(0.02)
MIL grades4	-0.127***	0.027	0.04*	0.031	0.008	0.017
	(0.011)	(0.027)	(0.022)	(0.02)	(0.021)	(0.018)
STEM grades1	-0.35***	0.296***	0.303***	0.316***	0.313***	0.255***
	(0.018)	(0.038)	(0.032)	(0.03)	(0.03)	(0.029)
STEM grades2	-0.523***	0.038	0.044	0.033	0.036	-0.021
	(0.018)	(0.036)	(0.031)	(0.029)	(0.029)	(0.028)
STEM grades3	-0.621***	-0.097**	-0.06	-0.06*	-0.063*	-0.108***
	(0.018)	(0.04)	(0.033)	(0.031)	(0.032)	(0.029)
STEM grades4	-0.588***	-0.113***	-0.09***	-0.091***	-0.083**	-0.129***
	(0.017)	(0.037)	(0.031)	(0.029)	(0.029)	(0.028)
MajorGrade3	-0.474***	-0.039	-0.061	-0.065*	-0.041	-0.11***
	(0.023)	(0.05)	(0.041)	(0.038)	(0.041)	(0.037)
MajorGrade4	-0.482***	-0.066	-0.039	-0.034	-0.073*	-0.083**
	(0.02)	(0.047)	(0.04)	(0.036)	(0.037)	(0.035)
Majorgrade5	-0.425***	-0.121***	-0.108***	-0.087**	-0.108***	-0.139***
	(0.019)	(0.041)	(0.035)	(0.032)	(0.033)	(0.031)
MajorGrade6	-0.43***	-0.116***	-0.103***	-0.089**	-0.119***	-0.13***
	(0.019)	(0.043)	(0.036)	(0.033)	(0.034)	(0.031)
Youngsterdrop	0.041***	-0.041**	-0.058***	-0.051***	-0.061***	-0.042***
	(0.007)	(0.019)	(0.015)	(0.013)	(0.015)	(0.012)
Plebedrop	0.005	-0.043***	-0.064***	-0.06***	-0.049***	-0.05***
	(0.005)	(0.014)	(0.011)	(0.01)	(0.011)	(0.008)
Major Switch	0.076***	0.111***	0.107***	0.098***	0.117***	0.097***
	(0.006)	(0.016)	(0.012)	(0.011)	(0.012)	(0.009)
Tutor	-0.015***	0.004	0.003	0.003	0.004*	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)
End Group 1	-0.151***	0.013	0.01	0.013	0.015	-0.001
	(0.009)	(0.018)	(0.016)	(0.015)	(0.015)	(0.015)
End Group 2	0.03***	0.058**	0.067***	0.064***	0.058***	0.056***
	(0.008)	(0.019)	(0.016)	(0.015)	(0.015)	(0.014)
End Group 3	0.113***	0.003	0.024	0.016	0.014	0.028
	(0.009)	(0.023)	(0.019)	(0.018)	(0.018)	(0.016)

Standard errors in parentheses. See Appendix 3 for variable descriptions. *** .01, **.05, * .1 significance.

ii. Foundation Schools

Trends in the Foundation school matched results mirror some of the results from matched results on the NAPS cohort. Table 4b illustrates the unmatched and matched comparison of effects of Foundation school treatment. The cohort of Foundation students graduate at a rate nearly 10% higher than those who do not attend a precollege program. The most modest estimate of positive returns to graduation rates, using the caliper method, generates an estimate that those who attend a Foundation school graduate at a rate 8.0% higher than students of comparable backgrounds who do not attend Foundation school. This is similar to the result gleaned from data on NAPS participants.

Unlike for the NAPS cohort, for the Foundation cohort returns to class rank are positive. Matched results suggest that overall order of merit is positively impacted by attendance at the Foundation program. However, these results are marginal and significant only for two of the five matching methods. Matching based on the caliper method suggests a 2.2% positive change in class rank due to attendance at a Foundation school. Kernel matching, on the other hand, suggests a 3.6% negative shift in class rank based on attendance at a Foundation school. Academic order of merit results are similarly confusing, and no results are statistically significant. Mirroring the positive returns to military performance grades, the military order of merit measure is positively impacted by attendance at a Foundation school in four out of the five matching methods. Military order of merit is increased by 4.5% using the most modest estimate, the caliper method.

In terms of academic coursework, results for the Foundation cohort mirror NAPS results. There are significant positive returns to the Foundation school program in the first semester of academic coursework. Even the most modest estimate using the 20 nearest neighbors suggests a

.134 increase in academic grade improvement. However, by the second semester these positive returns are diminishing and only significant in two out of the five matching methods. By the third semester, returns have become negative although they are still only significant in two out of the five matching methods. The magnitude of the negative returns is much lower than the magnitude of positive returns in the first semester, the most modest estimate based on 20 nearest neighbors is a difference of -.044. The fourth semester results are insignificant. The resulting magnitudes are comparable to the magnitude differences between the NAPS and direct cohort. In other words, these results are similar to NAPS results in both trend and magnitude.

Looking more specifically at positive returns to STEM grades from the Foundation school program, unmatched results suggest large magnitude negative returns in all four semesters. However, the matched results show a familiar trend. First semester positive returns to STEM grades are significant across the board except for using kernel matching. Even the most modest estimate suggests .136 positive difference in STEM grade point average using the caliper method of matching. By second semester, positive returns have diminished to insignificance, and by the third and fourth semesters there are significant negative returns to attending a Foundation school. The negative returns are significant across all methods of matching, and even the most modest for third semester is -.109 using the five nearest neighbors, and -.061 for the fourth semester using the caliper method. Comparing Foundation results to NAPS results, the magnitude of returns to STEM grades is halved in the first semester and nearly doubled in the third semester. In other words, the initial positive returns in the first semester are much smaller than returns to the NAPS program, and the negative effects in the third semester are much larger.

These matched results again suggest the troubling conclusion that Foundation schools are not only not helping, but due to some aspect of the pre-college cohort, the group performs significantly worse in STEM course grades than individuals with similar backgrounds who entered USNA directly from high school.

Explanations for these negative returns again include the possibility of peer effects having an impact on the academic performance of Foundation students. Another explanation is that by the third semester of academic coursework, students are taking so many different courses and have different majors that variance and noise have increased to the point that a comparison is not feasible.

The returns to military performance grades are surprising based on the trends seen in academic and STEM grades. Foundation students perform better militarily than comparable direct entry students in each semester across nearly all matching methods. Moreover, the magnitude of performance stays fairly consistent. First semester, kernel matching gives the most modest estimate of positive returns to be .086. Kernel matching is again the most modest at .072 for second semester and .075 third semester. By the fourth semester, positive gains from Foundation school have diminished, with the most modest estimate suggesting positive returns of .032 based on an analysis of 20 nearest neighbors.

The variables that describe the type of major from which a student graduates are broken down by USNA designation. The *End Group 1* and *End Group 2* variables are insignificant across the board. However, the *End Group 3* variable suggests positive returns to ending in a group 3 major if a students attends a Foundation school. As described above, group 3 majors include the humanities and social sciences. Specifically, the cohort of Foundation school students is, by the most modest estimate, 6.7% more likely to graduate as a group three major than the matched cohort of direct entry students. Effects on major course grades are statistically insignificant during both the sophomore and junior year.

Like NAPS students, Foundation school students are less likely to leave USNA before graduation. The *Youngsterdrop* and *Plebedrop* variables both indicate that the Foundation school program has a positive impact on retention during freshman and sophomore year. According to matched results, Foundation school students are between 4.8% and 6.8% less likely to drop out during their youngster year, and between 2.1% and 3.2% less likely to drop out during their freshman year. This may reflect the fact that Foundation school students feel the burden of preparatory school tuition. While NAPS is free for students, Foundation schools require families to contribute some tuition based on financial need. Those students who attended Foundation schools may feel extra pressure to graduate from USNA based on the fact that their families have contributed financial resources to their success. Also similar to the NAPS cohort results, matched results for Foundation students indicate that the treated cohort switch majors between 4.9 and 5.9 percentage points more often that direct entry students.

The conclusions about the Foundation school cohort of treated students echoes the conclusions about the returns for the NAPS program. Propensity score matching suggests positive returns to sending students to Foundation schools. Students perform better in their first semester coursework, are more likely to graduate, and less likely to drop out during their freshman and sophomore years. Foundation schools also appear to contribute to significant positive performance in the military aspect of USNA life. This makes sense based on the fact that many Foundation schools are military based preparatory school programs. As in the NAPS conclusions, positive returns to academic performance appear to dwindle into insignificance and eventually turn negative by the third and fourth semester. This suggests the disturbing possibility that negative returns to investment in preparatory school programs are not isolated to the USNA specific NAPS program, but that this result may be a symptom of programs nationwide.

Table 4b – Second Stage Matched Results Comparing Foundation and Direct Entry Students

	Unmatched	NN(1)	NN(5)	NN(20)	Caliper	Kernel
Graduation	0.058***	0.095***	0.096***	0.09***7	0.095***	0.08***
	(0.011)	(0.016)	(0.012)	(0.011)	(0.016)	(0.01)
OOM	0.109***	-0.022*	-0.015	-0.014	-0.022*	0.036***
	(0.008)	(0.011)	(0.009)	(0.008)	(0.011)	(0.008)
AOM	0.123***	-0.008	-0.003	0.0004	-0.009	0.048***
	(0.008)	(0.012)	(0.009)	(0.009)	(0.012)	(0.008)
MOM	0.045***	-0.046***	-0.047***	-0.047***	-0.045***	-0.01
	(0.008)	(0.012)	(0.01)	(0.009)	(0.012)	(0.009)
AC grades1	-0.155***	0.148***	0.143***	0.134***	0.148***	0.003
	(0.018)	(0.025)	(0.019)	(0.018)	(0.025)	(0.017)
AC grades2	-0.232***	0.042*	0.044**	0.035*	0.046	-0.084***
	(0.018)	(0.024)	(0.018)	(0.017)	(0.024)	(0.016)
AC grades3	-0.283***	-0.045	-0.054**	-0.044*	-0.045	-0.146***
	(0.019)	(0.027)	(0.02)	(0.019)	(0.025)	(0.018)
AC grades4	-0.256***	-0.035	-0.028**	-0.037*	-0.034	-0.128***
	(0.019)	(0.025)	(0.019)	(0.018)	(0.025)	(0.017)
MIL grades1	0.045***	0.103***	0.113***	0.115***	0.101***	0.086***
	(0.013)	(0.02)	(0.015)	(0.015)	(0.02)	(0.014)
MIL grades2	0.026**	0.098***	0.108***	0.104***	0.094***	0.072***
	(0.013)	(0.019)	(0.014)	(0.013)	(0.019)	(0.012)
MIL grades3	0.039**	0.085***	0.109***	0.098***	0.084***	0.075***
	(0.015)	(0.022)	(0.017)	(0.016)	(0.022)	(0.015)
MIL grades4	-0.041**	0.049**	0.035**	0.032*	0.047**	-0.001
	(0.014)	(0.02)	(0.015)	(0.014)	(0.02)	(0.014)
STEM grades1	-0.171***	0.139***	0.163***	0.178***	0.136***	0.022
	(0.023)	(0.032)	(0.025)	(0.023)	(0.032)	(0.022)
STEM grades2	-0.304***	0.027	0.043	0.027	0.027	-0.123***
	(0.023)	(0.032)	(0.025)	(0.023)	(0.032)	(0.022)
STEM grades3	-0.386***	-0.133***	-0.109***	-0.11***	-0.132***	-0.224***
	(0.023)	(0.032)	(0.025)	(0.023)	(0.032)	(0.022)
STEM grades4	-0.339***	-0.066**	-0.078***	-0.078***	-0.061*	-0.186***
	(0.022)	(0.03)	(0.023)	(0.022)	(0.03)	(0.021)
MajorGrade3	-0.21***	-0.012	-0.027	-0.017	-0.01	-0.094***
	(0.027)	(0.039)	(0.031)	(0.029)	(0.04)	(0.028)
MajorGrade4	-0.179***	0.03	0.01	0.008	0.039	-0.067**
	(0.025)	(0.036)	(0.028)	(0.026)	(0.036)	(0.025)
Majorgrade5	-0.193***	-0.027	-0.033	-0.031	-0.027	-0.092***
	(0.023)	(0.032)	(0.025)	(0.024)	(0.032)	(0.022)
MajorGrade6	-0.136***	0.035	0.013	0.027	0.034	-0.04
	(0.023)	(0.033)	(0.026)	(0.024)	(0.033)	(0.023)
Youngsterdrop	-0.026**	-0.061***	-0.069***	-0.067***	-0.068***	-0.048***
	(0.009)	(0.013)	(0.009)	(0.009)	(0.013)	(0.008)
Plebedrop	-0.007	-0.026**	-0.031***	-0.032***	-0.024**	-0.021***
	(0.006)	(0.009)	(0.007)	(0.006)	(0.009)	(0.006)
Major Switch	0.049***	0.049***	0.056***	0.059***	0.049***	0.054***
	(0.008)	(0.011)	(0.008)	(0.007)	(0.011)	(0.007)
Tutor	-0.014***	-0.006	-0.003	-0.002	-0.006	-0.007**
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)
End Group 1	-0.067***	0.019	0.008	0.01	0.017	-0.025*
	(0.012)	(0.017)	(0.013)	(0.012)	(0.017)	(0.012)
End Group 2	-0.013	0.01	0.006	0.006	0.009	-0.002
	(0.011)	(0.015)	(0.012)	(0.011)	(0.015)	(0.011)
End Group 3	0.138***	0.067***	0.084***	0.081***	0.069***	0.107***
	(0.012)	(0.019)	(0.015)	(0.014)	(0.019)	(0.013)

Standard errors in parentheses. See Appendix 3 for variable descriptions. *** .01, **.05, * .1 significance.

iii. Prior College

An analysis of educational returns for students who attended a prior year of college bear almost no comparison to the matched NAPS and Foundation school results. Students who attended college prior to USNA see positive returns to nearly all performance variables across nearly all methods of matching. Table 4c illustrates the unmatched and matched comparison of effects of college on performance at USNA. It is important to note that propensity score matching is used to compare students of similar backgrounds who are either assigned treatment or not assigned treatment. In this case, by virtue of rejection from the Academy, this cohort of students had the means and desire to enroll in another college and re-apply to USNA for the following year. Therefore, it is unrealistic to make an apples to apples comparison of this type of treatment with other “pre-college” treatments like NAPS and Foundation schools. However, it is informative to examine returns of a year of college in order to draw conclusions about the effectiveness of pre-college preparatory programs versus other education systems on a broader national scale.

Prior college students graduate at a higher rate than the similar cohort of direct entry students; however, the magnitude of the difference in graduation rates is not as dramatic as the comparison between Foundation and direct entry, or NAPS and direct entry groups. The kernel method is the most modest in estimating that prior college students graduate at a rate 3.7% higher than matched direct entry students. Matching based on the five nearest neighbors estimates a graduation rate for prior college students that is 6.8% higher than direct entry students.

Returns to class rank are also positive. Matched results suggest that overall order of merit is positively affected between 6.1% and 8.4% by a year of outside college. Academic order of merit results are significant on three out of the five matching methods and suggest higher rank by between 3.0% and 5.6%. The military order of merit measure is also significant on three out of the five matching methods and is positive between 3.6% and 6.0%.

In terms of academic coursework, results for the college cohort are large in magnitude and positive across all four semesters. First semester positive returns on academic grades range between .21 using the kernel method and .373 using 20 nearest neighbors. Second semester the most modest estimate using kernel matching is .132 and the highest magnitude match is .282 using 20 nearest neighbors. Third semester results range between .15 and .291, and fourth semester results range between .152 and .28. All four semesters have large magnitude positive returns to academic grades from a prior year of college.

Returns to STEM grades mirror the returns to academic grades, but with even more dramatic magnitudes. First semester positive returns are between .274 on the low end kernel estimate, and .526 matching on the nearest neighbor. This suggests that those students who attend a year of college are outperforming students of similar backgrounds who were admitted directly to USNA by over .5 on a 4.0 scale. In other words, their STEM grades are 12.5% higher than the comparable direct cohort. Second semester returns are between .155 and .347. Third semester returns are between .17 and .409, and fourth semester returns are between .144 and .295. Not only are returns to STEM grades significantly positive, but also the positive returns persist far past the first semester into the fourth semester, unlike the positive returns seen from the NAPS and Foundation programs.

These matched results suggest that having a year of college before attending USNA can provide students not only with the course familiarity to succeed in their freshmen courses, but it also equips them with other skills that cause them to outperform direct entry students of similar backgrounds in later semesters when coursework is less standard and is more unfamiliar. These results beg the question of why we do not simply send unprepared students to a year of community college or other college and do away with the NAPS and Foundation school programs.

The returns to military performance grades are again positive and significant across all four semesters and in all five methods of matching. First semester returns range from positive .185 to .213. By second semester the magnitude has decreased to a range between .052 and .101. However, by the third semester, the range has increased again to between .102 and .138. By the fourth semester, returns to military performance have increased again to between .138 and .213. These results are not only positive, but the magnitude is higher than the consistent positive returns to Foundation school programs.

There are marginally significant returns to graduating as a group 1 major or a group 3 major. Results on the *End Group 2* variable are insignificant across the board. The *End Group 1* variable suggests positive returns to ending in a group 1 major if a students attends a year of college before matriculating at USNA. As described above, group 1 majors include only the engineering majors at USNA. Specifically, the cohort of college students is, by the most modest estimate, 4.4% more likely to graduate as a group one major than the matched cohort of direct entry students. However, results are only significant for three of the five matching methods. Similarly, although results are only significant for two of the five matching methods, college students are, on the low end, 4.5% more likely to graduate as humanities or social science majors than their direct entry counterparts.

Interestingly, retention rates for freshman and sophomore year are less significant and of lower magnitude than retention rates for the NAPS and Foundation cohorts. The *Younsterdrop* variable is only significant on two out of the five matching methods, and the college cohort is only between 3.5% and 4.0% less likely to attrite during sophomore year. Attrition rates for freshmen year are insignificant for all matching methods. Finally, unlike the NAPS and Foundation cohort, college students are significantly more likely to tutor their peers, according to the *Tutor* variable. Returns suggest that the cohort is between 2.3% and 3.1% more likely to be a student tutor than the comparable cohort of direct entry students.

Propensity score matching suggests quantitative positive returns to having a year of college experience before attending USNA. In some ways, this conclusion seems obvious: after a year of college, a student will be more successful at managing the college lifestyle. However, these results also seem to condemn the practice of sending students to a year of preparatory school. If the cohort of students who attend a year of college before USNA perform significantly better than their comparable counterparts in all aspects of academy life, should the Academy just send all unprepared students to a year of college or community college before coming to USNA, and do away with preparatory school programs? Boiled down, these results suggest that a year at college does significantly more for a student's long-term performance in college than preparatory schools do. However, it is important to consider the particular cohort of students who make up the college treatment group because of the selectivity issue inherent to interpreting these results. These are students who had the resources and ability to get into and attend another college instead of USNA for one year. They are also a group of people with the drive and motivation to go through a second freshman year at USNA after first attending another school. This separates them from the cohort of students who are sent to NAPS and Foundation schools.

Table 4c – Second Stage Matched Results Comparing College and Direct Entry Students

	Unmatched	NN(1)	NN(5)	NN(20)	Caliper	Kernel
Graduation	0.036*	0.061**	0.068***	0.06***	0.061**	0.037*
	(0.018)	(0.025)	(0.018)	(0.017)	(0.025)	(0.016)
OOM	-0.019	-0.061***	-0.084***	-0.087***	-0.062***	-0.022
	(0.013)	(0.018)	(0.015)	(0.014)	(0.018)	(0.013)
AOM	0.012	-0.03	-0.05***	-0.056***	-0.03	0.009
	(0.013)	(0.019)	(0.015)	(0.014)	(0.019)	(0.013)
MOM	-0.014	-0.036	-0.057***	-0.06***	-0.036	-0.016
	(0.014)	(0.019)	(0.015)	(0.014)	(0.019)	(0.013)
AC grades1	0.205***	0.332***	0.363***	0.373***	0.33***	0.21***
	(0.031)	(0.044)	(0.034)	(0.032)	(0.044)	(0.031)
AC grades2	0.128***	0.257***	0.277***	0.282***	0.258***	0.132***
	(0.029)	(0.042)	(0.033)	(0.031)	(0.042)	(0.03)
AC grades3	0.144***	0.288***	0.278***	0.291***	0.29***	0.15***
	(0.031)	(0.045)	(0.035)	(0.033)	(0.046)	(0.032)
AC grades4	0.147***	0.272***	0.278***	0.28***	0.273***	0.152***
	(0.03)	(0.042)	(0.031)	(0.029)	(0.042)	(0.028)
MIL grades1	0.184***	0.207***	0.21***	0.213***	0.208***	0.185***
	(0.022)	(0.031)	(0.024)	(0.023)	(0.031)	(0.022)
MIL grades2	0.051**	0.089***	0.101***	0.099***	0.091*	0.052*
	(0.021)	(0.029)	(0.023)	(0.021)	(0.029)	(0.021)
MIL grades3	0.101***	0.122***	0.124***	0.138***	0.124***	0.102***
	(0.024)	(0.034)	(0.026)	(0.025)	(0.034)	(0.024)
MIL grades4	0.136***	0.21***	0.187***	0.175***	0.213***	0.138***
	(0.022)	(0.032)	(0.026)	(0.024)	(0.032)	(0.024)
STEM grades1	0.269***	0.526***	0.474***	0.46***	0.522***	0.274***
	(0.038)	(0.056)	(0.042)	(0.039)	(0.056)	(0.038)
STEM grades2	0.149***	0.346***	0.315***	0.307***	0.347***	0.155***
	(0.038)	(0.054)	(0.045)	(0.042)	(0.054)	(0.041)
STEM grades3	0.164***	0.406***	0.304***	0.329***	0.409***	0.17***
	(0.038)	(0.053)	(0.041)	(0.039)	(0.053)	(0.038)
STEM grades4	0.137***	0.288***	0.294***	0.295***	0.29***	0.144***
	(0.036)	(0.05)	(0.038)	(0.036)	(0.05)	(0.034)
MajorGrade3	0.147***	0.276***	0.241***	0.247***	0.276***	0.151***
	(0.043)	(0.058)	(0.045)	(0.042)	(0.058)	(0.041)
MajorGrade4	0.104***	0.254***	0.233***	0.218***	0.254***	0.11***
	(0.038)	(0.052)	(0.039)	(0.036)	(0.052)	(0.035)
Majorgrade5	0.083**	0.19***	0.169***	0.182***	0.192***	0.088**
	(0.036)	(0.047)	(0.037)	(0.035)	(0.047)	(0.034)
MajorGrade6	0.04	0.078	0.135***	0.138***	0.078***	0.044
	(0.036)	(0.05)	(0.039)	(0.036)	(0.05)	(0.035)
Youngsterdrop	-0.022	-0.027	-0.035**	-0.04**	-0.027	-0.022
	(0.014)	(0.019)	(0.015)	(0.013)	(0.019)	(0.013)
Plebedrop	-0.003	-0.008	-0.018	-0.015	-0.008	-0.003
	(0.01)	(0.014)	(0.011)	(0.01)	(0.014)	(0.01)
Major Switch	0.005	-0.004	0.008	0.011	-0.004	0.006
	(0.013)	(0.018)	(0.014)	(0.013)	(0.018)	(0.013)
Tutor	0.026***	0.023**	0.027**	0.031***	0.023*	0.026*
	(0.006)	(0.011)	(0.01)	(0.009)	(0.011)	(0.009)
End Group 1	0.003	0.059**	0.044*	0.039	0.059**	0.004
	(0.02)	(0.027)	(0.022)	(0.021)	(0.028)	(0.02)
End Group 2	-0.023	-0.036	-0.022	-0.012	-0.04	-0.023
	(0.017)	(0.024)	(0.018)	(0.017)	(0.024)	(0.016)
End Group 3	0.057**	0.038	0.045**	0.033	0.041	0.056**
	(0.019)	(0.028)	(0.022)	(0.021)	(0.028)	(0.02)

Standard errors in parentheses. See Appendix 3 for variable descriptions. *** .01, **.05, * .1 significance.

D. Instrumental Variable Regression

The first specification utilizing the IV includes the same first stage background characteristics used in the propensity score matching specification: *Math SATs, Verbal SATs, high school rank, and high school quality*. In addition, the first specification includes *gender*, ethnicity dummy variables, and state dummy variables. The second stage specification includes only the four main variables: *Math SATs, Verbal SATs, high school rank, and high school quality*.

We regressed six performance variables as dependent variables in order to analyze the effect of NAPS on USNA performance. We examined the first two semesters of academic grades, the binary variable indicating graduation, and the normalized academic order of merit. These performance variables provide a snapshot with which to compare propensity score matching and OLS.

The IV regression on *AcGrade1* and *AcGrade2* indicates positive returns to the first semester at USNA, with NAPS participation adding .604 in academic grade point average. However, by the second semester, returns have diminished to insignificance. The graduation logistic results indicate a similar result as the propensity score matching: NAPS participants graduate 1.43 more than non participants with comparably the same background characteristics. Finally, similar to the PSM results, academic order of merit is negatively impacted by 3.6 percentage points. Table 5a indicates coefficients on the IV and the contribution of the four main variables to the model.

Table 5a – IV Regression Using Propensity Score As Instrumental Variable – Specification 1

	Ac Grade 1	Ac Grade 2	Graduation Rate (Logistic)	Grad AOM
IV - Pscore	0.604***	-0.0001	1.434**	0.036**
	(0.053)	(0.001)	(0.201)	(0.0157)
Highest Math SAT	0.0038***	-0.000001	1.004***	-0.001***
	(0.0001)	(0.000003)	(0.0004)	(0.00004)
Highest Verbal SAT	0.002***	0.000001	0.9998	-0.0008**
	(0.0001)	(0.000003)	(0.0003)	(0.00004)
High School Rank (Percentile)	0.119	-0.0007	1.57*	-0.033
	(0.086)	(0.002)	(0.329)	(0.025)
High School Quality Measure	0.002***	0.000002	1	-0.0008***
	(0.00009)	(0.000002)	(0.0002)	(0.00003)
Number of Observations	21053	21117	21053	16533
R-squared	0.119	0.0001	Pseudo: 0.011	0.3219
Adjusted R-squared	0.1188	-0.0002	--	0.3217

Standard errors in parentheses. See Appendix 3 for variable descriptions. *** .01, **.05, * .1 significance.

The second specification includes dummy variables by year as well as by state. This more inclusive model has similar results as the one above, but with nearly doubly the magnitudes on the coefficient of the IV. Academic grade impact diminishes from .9906 in the first semester to insignificance by the second semester. Graduation rates are improved by 3.44 times for NAPS participants versus non participants. Finally, academic order of merit is decreased by 6.9 percentage points. Table 5b indicates the coefficients on the second stage, second specification IV model.

Table 5b – IV Regression Using Propensity Score As Instrumental Variable – Specification 2

	Ac Grade 1	Ac Grade 2	Graduation Rate (Logistic)	Grad AOM
IV - Pscore	0.9906***	-0.0001	3.442***	0.069***
	(0.0513)	(0.0013)	(0.48)	(0.015)
Highest Math SAT	0.0043***	-0.000001	1.005***	-0.001***
	(0.0001)	(0.000003)	(0.0004)	(0.00004)
Highest Verbal SAT	0.0022***	0.0000005	1.001***	-0.0007***
	(0.0001)	(0.000003)	(0.0003)	(0.00004)
High School Rank (Percentile)	0.3269***	-0.001	2.457***	-0.016
	(0.0856)	(0.002)	(0.578)	(0.025)
High School Quality Measure	0.0017***	0.000002	1.001***	-0.0008***
	(0.0001)	(0.000002)	(0.0003)	(0.00003)
Number of Observations	21053	21053	21053	16533
R-squared	0.1291	0.0001	Pseudo: 0.0145	0.3225
Adjusted R-squared	0.1289	-0.0002	--	0.3223

Standard errors in parentheses. See Appendix 3 for variable descriptions. *** .01, **.05, * .1 significance.

VII. Conclusions

The current literature in this field suggests several implications of pre-college education on college performance. First, college graduation rates are declining as a result of lack of student preparation. In contrast to the findings of Adleman (1999) and Attewell, Lavin, Domina, and Levey (2006), results from this study suggest that pre-college education can significantly improve college retention rates and increase graduation rates. Specifically, graduation rates for NAPS and Foundation students were between 7% and 10% higher than for comparable direct admissions students. In terms of degree attainment, pre-college education improves a student’s likelihood of making it through a four-year selective institution with a degree. As discussed in the body of this paper, there are several possible reasons for this result. Echoing the hypothesis of the Soliday (2002) paper, it is possible that the impact of pre-college programs like NAPS and Foundation schools is not quantified in the form of higher grades but does contribute to the development of skills which help students persevere where similar students without pre-college experience drop out before attaining a degree. Another explanation for generalizations about graduation rates is the idea that pre-college students feel added pressure to attain a degree due to their investment in the form of time have and, in the case of Foundation students, financial resources to their education.

This study makes a unique contribution to current literature through analysis of performance semester by semester over four years. However, results suggest troubling conclusions about how preparatory school influences a student’s academic performance in the classroom. Although returns from preparatory school programs are highly positive in the first semester, the trend is that these positive returns diminish rather than persist through a student’s four years at college. In fact, propensity score matching techniques suggest that the significance of the positive returns stops after the first semester, and diminishes to the point that returns from pre-college are negative. This generalization has several implications. First, preparatory school programs are preparing students for specific introductory level courses rather than giving them study skills and familiarity with college level work. Second, there is too much noise after the first semester, to assess the impact of remedial education in a meaningful way. The third potential conclusion is that remedial education participants respond to peer effects that create a cohort in which educational persistence is lauded, while academic performance is undervalued. This would explain how the cohort of pre-college participants actually drag each other down academically, leading to negative returns to academic grades in the third and fourth semesters. This would also help explain the negative returns to academic order of merit that span all forms of pre-college education other than participation in prior college.

As discussed in the body of this work, the NAPS program is similar to many national preparatory programs. Moreover, the inclusion of data on Foundation school students allows us to extrapolate conclusions to a national scale. Unfortunately, this suggests the disturbing conclusion that investing in pre-college programs is not a viable method for improving human capital attainment for the U.S. workforce.

Pre-college education plays an important role in student development and particularly in student persistence toward earning a degree. However, the following question can be raised: is the cost of these remedial type programs worth the increase in graduation rates at the expense of academic performance? Is the goal of selective institutions to push students toward a degree at all costs or to actually improve educational achievement through pre-college programs? These are questions that must be considered to accurately assess the cost versus benefit of pre-college education.

VIII. Bibliography

Adelman, C. (1999). Answers in the toolbox: Academic intensity, attendance patterns, and bachelor's degree attainment. Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education. Retrieved from http://www.ed.gov/pubs/Toolbox/toolbox.html

Alliance for Excellent Education. "Saving Now and Saving Later: How High School Reform Can Reduce the Nation's Wasted Remediation Dollars." 2011. www.all4ed.org.

Attewell, P., Lavin, D., Domina, T., and Levey, T. (2006). "New Evidence on College Remediation." Journal of Higher Education, Vol. 77, No. 5 (September/October 2006).

Bettinger, Eric and Bridget Long. (2004). "Shape Up or Ship Out: The Effects of Remediation on Students at Four-Year Colleges." NBER Working Paper 10369. www.nber.org/papers/w10369.

Bettinger, Eric and Bridget Long. (2005). "Addressing the Needs of Under-Prepared Students in Higher Education: Does College Remediation Work?" Journal of Human Resources, 44(3):.

Boarding School Review. "Post-Graduate Boarding Schools." Boarding School Review LLC. http://www.boardingschoolreview.com/pg_boarding_schools.php. April 21, 2013.

Bound, John, Michael Lovenheim and Sarah Turner. 2009. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources." American Economic Journal: Applied Economics, 2(3): 129-57.

Breneman, David and William Haarlow. 1997. "Remedial Education: Costs and Consequences." Remediation in Higher Education. Washington D.C.: Thomas B. Fordham Foundation.

Carnelvale, A., Smith, N. and Jeff Strohl. "Help Wanted: Projections of Jobs and Education Requirements through 2018." Washington, D.C.: Georgetown University Center on Education and the Workforce, 2010.

Carrell, S., Maghakian, T., and West, J. (2010). "A's from Zzzz's? The Causal Effect of School Start Time on the Academic Acheivment of Adolescents." American Economic Journal 3(3): 62-71.

Carrell, Scott, Richard Fullerton and James West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." Journal of Labor Economics, 27(3): 439-464.

Doyle, William R. 2007. "The Effect of Community College Enrollment on Bachelor's Degree Completion." Vanderbilt University working paper.

Ortman, A., Slobodyan, S. & Nordberg, S. 2003. "(The Evolution of) Postsecondary Education: A Computational Model and Experiments." mimeo.

Green, Jay and Greg Foster. 2003. "Public High School Graduation and College Readiness Rates in the United States." Manhattan Institute, Center for Civic Information, Education Working paper no. 3.

Heinrich, Carolyn, Alessandro Maffioli, and Gonzalo Vazquez. 2010. "A Primer for Applying Propensity-Score Matching." Office of Strategic Planning and Development Effectiveness, Inter-American Development Bank, Technical Notes No. IDB-TN-161.

Hollenbeck, Kevin. 1992. "Postsecondary Education as Triage: Returns to Academic and Technical Programs." Upjohn Institute Working Paper No. 92-10. Kalamazoo, Mich.: W.E. Upjohn Institute for Employment Research.

Kane, Thomas J. and Cecelia Rouse. 1999. "The Community College: Educating Students at the Margin between College and Work." Journal of Economic Perspectives, 13(1): 63-84.

Lavin, D., Alba, R., and Silberstein, R. (1981). "Right versus privilege: The open admissions experiement at the City University of New York." New York: Free Press.

Lemieux, Thomas. 2006. "Postsecondary Education and Increasing Wage Inequality." American Economic Review, 96(2): 195–199.

Long, Bridget Terry and Michael Kurlaender. forthcoming. "Do Community Colleges Provide a Viable Pathway to a Baccalaureate Degree?" Educational Evaluation and Policy Analysis.

Matsudaira, Jordan D. 2007. "Mandatory Summer School and Student Achievement: A Regression Discontinuity Analysis." Journal of Econometrics, 142(2): 829-850.

Merisotis, J., and Phipps, R. (2000). "Remedial Education in Colleges and Universities: What's Really Going On?" The Review of Higher Education, 24(1), 67-85.

Polacheck, Solomon. 2004. "How the Human Capital Model Explains Why the Gender WageGap Narrowed." Institute for the Study of Labor Discussion Paper Series. April 2004.

Regan, Tracy L., Ronald L. Oaxaca and Galen Burghardt. 2009. "A Human Capital Model of the Effects of Ability and Family Background on Optimal Schooling Levels." University of Miami, Department of Economics. Working Papers. 2010-8.

Rosenbaum, P., and D. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika, 70(1): 41-55.

Rouse, Cecilia. 1995. "Democratization or Diversion? The Effect of Community Colleges on Educational Achievement." Journal of Business and Economic Statistics, vol. 13, no. 2, April, pp. 217-24.

Schneider, Mark and Lu Yin. 2011. "The Hidden Costs of Community Colleges." American Institutes for Research.

Soliday, Mary. 2002. The Politics of Remediation. Pittsburgh: The University of Pittsburgh Press.

United States Naval Academy. www.USNA.edu. Accessed 20MAR13.

Van der Klaauw, Wibert. 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach." International Economic Review, 43(4): 1249-1287.

IX. Appendix 1 – Institutional Specifics

This paper takes into account the inherent selection bias that accompanies assignment to a remedial program. However, in this paper we examine pre-college remediation as a method for better preparing high school graduates for their college courses. Although USNA does have a few remedial courses for students who are unable to keep up in freshman level courses, as well as some students who validate freshman requirements, the vast majority of students take the same array of courses during their freshman year. Rather than examining remedial coursework undertaken when students are already enrolled at USNA, this study views each feeder program as a type of remediation, to which students are assigned based on background characteristics. Remediation is assigned as a condition of future appointment to USNA, eliminating the bias associated with optional remediation.

A. Background on the Naval Academy Preparatory School (NAPS)

The primary pre-college program of interest in this study, NAPS, was founded in 1915 to assist enlisted sailors in making the transition from military to academic life. Today, NAPS candidates receive standard benefits as active duty service members, including healthcare, standard pay, room, and board. Tuition for NAPS candidates is fully paid by the Department of the Navy. Finally, NAPS candidates are ensured matriculation into USNA barring any criminal or other serious offences. The average class at USNA is made up of over 15% NAPS graduates, with the 2011 class being comprised of over 17% NAPS graduates. Over the years, the mission of NAPS and the program's selection criteria have gradually evolved. Today the institution's mission is to "prepare selected candidates morally, mentally, and physically, with emphasis on strengthening the academic foundation of individual candidates for officer accession through the

U.S. Naval Academy" (www.USNA.edu). That is, the goal of NAPS is to provide potential USNA candidates (from either high school or the fleet) with the necessary academic skills to succeed in future college endeavors, thus providing positive returns to their human capital.

The NAPS program itself is run just like many of the top preparatory schools around the United States. NAPS students must take courses in English, Math, Chemistry, and Physics. Students are given an assessment at the beginning of the academic year and placed in courses according to their current level in each subject. An Academic Dean oversees the academic curriculum. A commanding officer, an executive officer, three company officers and two senior enlisted leaders run the school. Similar to preparatory schools around the country, NAPS students are housed in a dorm-like barracks building in Newport, Rhode Island. This environment away from home gives students a taste of the independence they will feel again as college freshmen. This aspect of preparatory school life is uniquely important to the development of self-motivated and driven students with the capacity to succeed as independent people away from home.

B. Background on the Naval Academy Preparatory School (NAPS)

Students who do not attend NAPS and are not admitted directly to USNA may be offered appointment to a different type of feeder program known as a Foundation school. USNA provides selected students the opportunity to enroll in other pre-college programs known as "Foundation" schools. Foundation students have up to 60% of their tuition covered by the Foundation Program and have a 95% guarantee of transfer to the Academy. Families are expected to contribute resources to pay for at least 40% of tuition based on their income level. Students offered the opportunity to go to Foundation school have the option to apply to 15 civilian and 4 military preparatory schools including Hargrave Military Academy, the Kent School, Peddie, and various other civilian and military preparatory schools (www.USNA.edu). A full list of current Foundation preparatory schools is included at the end of this section.

This program differs significantly from the NAPS program. Rather than being offered a place at a specific institution like NAPS, Foundation school students may choose their desired preparatory school, introducing variation in location, cost, and quality of education. Foundation school students make up a smaller cohort of precollege entrants at 6% of each entering USNA class. The Foundation program provides a different perspective on returns to investment in pre-college education program.

Similar to the NAPS program, the Foundation school program is designed as a preparatory year for students who are academically unprepared for USNA. Foundation students participate in what is known as a post-graduate or “PG” year. At preparatory school, they enroll in courses that include at minimum English, science, and math. Foundation schools fall into one of three categories: military preparatory schools, preparatory schools affiliated with a college or university, or independent preparatory schools.

Military preparatory schools, like the New Mexico Military Institute, are the most similar to the NAPS environment due to their militaristic nature. At military preparatory schools, students wear uniforms, participate in JROTC, and are organized in a hierarchical structure led by the most senior cadets.

Preparatory schools affiliated with a college or university, like Greystone Preparatory, boast access to the academic resources of a degree-granting institution. At Greystone, students are able to enroll in “advanced placement” or college level courses during their PG year in order to better prepare for college.

Finally, independent preparatory schools like the Salisbury School, boast a history and tradition of accepting students for a PG year in order to improve their academic preparedness and allow them an extra year as high school athletes to improve the likelihood of being recruited to a college for athletics.

The range of schools that participate in the Naval Academy Foundation program includes schools across all of the US. They are a cross-section of typical preparatory school programs. Like other preparatory schools, Foundation schools place high emphasis on academic performance, athletic performance, and college matriculation. Taken together, conclusions from the treatment groups described above can be extrapolated to broader national education strategy due to the unique characteristics of the USNA dataset.

Naval Academy Foundation Schools

1. Avon Old Farms School, Avon, Connecticut
2. Blair Academy, Blairstown, New Jersey
3. Greystone Preparatory School at Schreiner University, Kerrville, Texas
4. Hargrave Military Academy, Chatham, Virginia
5. The Hill School, Pottstown, Pennsylvania
6. The Hun School of Princeton, Princeton, New Jersey
7. Kent School, Kent, Connecticut
8. The Kiski School, Saltsburg, Pennsylvania
9. The Marion Military Institute, Marion, Alabama
10. The Mercersburg Academy, Mercersburg, Pennsylvania
11. New Mexico Military Institute, Roswell, New Mexico
12. Northfield Mount Hermon School, Northfield, Massachusetts
13. Northwestern Preparatory School, Crestline, California
14. The Peddie School, Hightstown, New Jersey
15. Portsmouth Abbey School, Portsmouth, Rhode Island
16. Salisbury School, Salisbury, Connecticut
17. Valley Forge Military Junior College, Wayne, Pennsylvania
18. Western Reserve Academy, Hudson, Ohio
19. Wyoming Seminary, Kingston, Pennsylvania

X. Appendix 2 – United States Naval Academy Majors by Group

Group 1 – Engineering and Weapons:
Aerospace Engineering
Computer Engineering
Electrical Engineering
General Engineering
Mechanical Engineering
Naval Architecture
Ocean Engineering
Systems Engineering

Group 2 – Mathematics and Science:
Chemistry
Computer Science
Cyber Operations
General Science
Information Technology
Mathematics
Oceanography
Operations Research
Physics
Quantitative Economics

Group 3 – Humanities and Social Sciences:
Arabic
Chinese
Economics
English
History
Political Science

XI. Appendix 3 – Variable Descriptions

Background Characteristics

African American	Binary variable indication whether a student classifies themselves as African American
Asian American	Binary variable indication whether a student classifies themselves as Asian American
Caucasian	Binary variable indication whether a student classifies themselves as Caucasian
Hispanic	Binary variable indication whether a student classifies themselves as Hispanic
Age on IDay	A student's age on their first day at the Naval Academy, Induction Day
Central	A binary variable indicating whether a student is from Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, or Wisconsin
Northern	A binary variable indicating whether a student is from Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, or Vermont
Pacific	A binary variable indicating whether a student is from Alaska, Arizona, California, Hawaii, Nevada, Oregon, Utah, or Washington
Southern	A binary variable indicating whether a student is from Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, or West Virginia
Western	A binary variable indicating whether a student is from Colorado, Idaho, Kansas, Montana, Nebraska, New Mexico, North Dakota, Oklahoma, South Dakota, Texas, or Wyoming
Verbal SAT	A student's highest reported SAT score for the verbal section
Math SAT	A student's highest reported SAT score for the math section
Sex	A binary variable indicating gender where 1=female
High School Quality Measure	A measure of the academic quality of individual high schools on the same scale as the SAT: from 200 to 800
High School Rank (Percent)	A student's percent rank in their high school class where .99 signifies being in the top 1% of high school graduating class
Military Father	Binary variable indicating whether the student's father was in the military
Military Mother	Binary variable indicating whether the student's mother was in the military

Performance Variables

AC grades1	First semester academic grade point average
AC grades2	Second semester academic grade point average
AC grades3	Third semester academic grade point average
AC grades4	Fourth semester academic grade point average
Academic Average	Average course GPA in all academic courses for the first four semesters excluding professional and military coursework
AOM	Normalized class rank based on eight semesters of academic grades
Graduated (indicator)	A binary variable indicating whether a student graduated
Major Switch	A binary variable indicating whether a student changed their major while at USNA
MajorGrade3	Third semester majors courses grade point average
MajorGrade4	Fourth semester majors courses grade point average
Majorgrade5	Fifth semester majors courses grade point average
MajorGrade6	Sixth semester majors courses grade point average
MIL grades1	First semester military grade point average
MIL grades2	Second semester military grade point average
MIL grades3	Third semester military grade point average
MIL grades4	Fourth semester military grade point average
Military Average	Average course GPA in all military performance grades for the first four semesters excluding professional and military coursework
MOM	Normalized class rank based on eight semesters of military performance grades
OOM	Normalized class rank based on combined AOM and MOM, eight semesters of academic grades and military performance grades
Plebedrop	A binary variable indicating whether a student left USNA during their freshman year
Professional Average	Average course GPA in all professional course grades for the first four semesters excluding professional and military coursework
Start Group 1	A binary variable indicating whether a student elected a major in group 1 during their freshman year
Start Group 2	A binary variable indicating whether a student elected a major in group 2 during their freshman year
Start Group 3	A binary variable indicating whether a student elected a major in group 3 during their freshman year
STEM grades1	First semester STEM grade point average
STEM grades2	Second semester STEM grade point average
STEM grades3	Third semester STEM grade point average
STEM grades4	Fourth semester STEM grade point average
Tutor	A binary variable indicating whether a student was involved in the student-tutor program while at USNA
Varsity Athlete (indicator)	A binary variable indicating whether a student was a varsity athlete while at USNA
Youngsterdrop	A binary variable indicating whether a student left USNA during their sophomore year

XII. Human Research Protection Program Approvals

U.S. Naval Academy Human Research Protection Program
Nimitz Library G10 - Mail Stop 10M - Annapolis, Maryland 21402

MEMORANDUM 14 December 2011

From: Ms. Erin Johnson, Academy's HRPP Office

To: MIDN 2/C Phoebe Kotlikoff, Economics Department

Subject: APPROVAL OF HUMAN SUBJECT RESEARCH

Ref: (a) SECNAVINST 3900.39D
(b) 32 CFR 219
(c) USNA HRPP Policy Manual

USNA Assurance # DoD N-40052 HRPP Approval # **USNA.2012.0006-IR-EM4-A**

1. The Superintendent, as the Institutional Official (IO), reviewed and approved your research protocol "The Effects of Post-Secondary Education on Midshipman Success at the United States Naval Academy" involving human subjects. The co-investigators are Assoc Prof Katherine A. Smith, Asst Prof Ahmed S. Rahman and MIDN 2/C Edward Butler from the Economics Department. It was determined to be exempt under 32 CFR 219.101(b)(4).

2. Research which is determined to be exempt under 32 CFR 219.101 is exempt from all regulatory requirements, unless there is a substantive change that could potentially alter the assessment of the exempt status. If there is a substantive change you must submit an amendment to your protocol in sufficient time to process the revisions and secure approval from the Superintendent. On an annual basis, a status update of all exempt studies will be conducted.

3. We would appreciate a notification of closure when the research has concluded according to Section XIII of the USNA HRPP Policy and Procedures manual and to provide this office with copies of any articles or presentations resulting from this research. Additionally, any presentations or publications must include acknowledgement of IRB approval using the HRPP approval number.

4. If you have any questions, please contact this office at 410-293-2533 or HRPPoffice@usna.edu.

ERIN JOHNSON
Academy's HRPP Office

DEPARTMENT OF THE NAVY
UNITED STATES NAVAL ACADEMY
121 BLAKE ROAD
ANNAPOLIS MARYLAND 21402-1300

3900
6 December 2011

MEMORANDUM

From: Chair, Institutional Review Board (Code 28)
To: Superintendent, United States Naval Academy

Subj: HUMAN SUBJECT RESEARCH BY MIDN 2/C PHOEBE KOTLIKOFF (ECONOMICS DEPARTMENT)

Ref: (a) SECNAVINST 3900.39D
(b) 32 CFR 219
(c) USNA HRPP Policy Manual

Encl: (1) Protocol Package for MIDN 2/C Phoebe Kotlikoff (Form 2, 3, 4, 5, CITI and Supplemental Information)

1. I have reviewed the research protocol submitted by MIDN 2/C Phoebe Kotlikoff from the Economics Department on "The Effects of Post-Secondary Education on Midshipman Success at the United States Naval Academy." Co-investigators are Asst Prof Ahmed S. Rahman, Assoc Prof Katherine A. Smith, and MIDN 2/C Edward Butler from the Economics Department.

2. This project evaluates the effects of attending a post-secondary education program on student performance at undergraduate institutions by using data from the USNA. Through regression analysis, the project will empirically demonstrate potential relationships between the level of participation in pre-college programs and success in undergraduate studies. De-identified data will be obtained from Institutional Research.

3. This research is determined to be exempt under 32 CFR 219.101(b)(4). Research which is determined to be exempt under 32 CFR 219.101 is exempt from all regulatory requirements, that includes continuing review, unless there is a substantive change that could potentially alter the assessment of the exempt status.

12/6/11

JUDITHANN HARTMAN

Date: 12-11-11

☑ Approved as recommended ☐ Conditionally Approved ☐ Disapproved

Comments:

M. H. MILLER
Vice Admiral, U.S. Navy
Superintendent